

Japan's researchers merit further reforms

The new batch of changes to procedures for selecting competitive grants is one of many recent welcome improvements to Japan's university research system. But it is by no means adequate.

JAPAN'S university researchers have good cause to be more optimistic about the future. After years of neglect, their government has in the past year introduced substantial new sources of research funds from ministries and agencies other than the Ministry of Education, Science, Sports and Culture (Monbusho), breaking that ministry's stranglehold on university research. On top of this, Monbusho itself has almost doubled the budget for its competitive research grants in the past five years bringing this source of funds up to a much more healthy ¥100 billion (about US\$1 billion) a year. And now Monbusho's Science Council, comprising leading academics, has introduced some much needed reforms (see page 573) to make the ministry's system of grant selection more open, fair and transparent.

This extra funding and the reforms across the whole system of government support for research are indeed welcome, and there can be little doubt that the changes will boost the output and quality of Japan's basic research. Nevertheless, much still remains to be done to improve the lot of Japan's university researchers.

One glaring problem that needs to be tackled is the chronic lack of technicians in the universities. Many years ago, the Japanese government decided to follow the lead of governments in the West and started to reduce the number of government employees. Smaller government was rightly seen as more efficient and as one way of tackling the swelling sea of red ink in government finances. But it has had disastrous consequences for university research.

Each year, the universities and Monbusho's university-related research institutes have to reduce their number of employees by a fixed percentage. This has largely been achieved by cutting technical rather than academic positions. Further exacerbating the problem are Monbusho's rigid rules on the salaries that can be paid to technicians. Technicians have always been treated as second-class citizens in Japan's universities and salaries are not competitive with those in industry. As a result, universities cannot attract high-quality technicians to the few posts still available. The decline in technical support has meant that an increasing burden of menial and technical tasks has fallen on young researchers, distracting them from creative research. Furthermore, expensive equipment is inadequately maintained, and safety is often compromised.

Those pushing for reform of the university system have some ideas about how the technician problem could be solved. One worthy of consideration is a proposal from Akito Arima, former president of Tokyo University, that large numbers of research assistant (*joshu*) positions, the lowest tenured academic rank in the university system, should be given to technicians, and the employment needs of young researchers should instead be met with more post-doctoral positions. With the government now planning to double the number of postdoctoral positions to 10,000 by the end of the

decade, Arima's idea is becoming a practical proposition.

The reforms of the ministry's competitive grant system are another step in the right direction that should provide young researchers with a better chance of success in their grant applications. But here too, much could still be done. One obvious improvement would be the provision of feedback on rejected grant applications. The present system, even with the reforms (which will be implemented soon), can be likened to a black hole. Applications go into the system but nothing comes out except a 'yes' or 'no'. As the majority of applications fail, this is demoralizing for researchers who do not know why their application failed or how close they came to the mark. The ministry's excuse that there are too many applications to provide feedback is lame, particularly when the Internet is there to allow information to be readily forwarded by e-mail.

Some exposure of Monbusho's grant screening process to scrutiny by non-Japanese scientists, as some are suggesting, would undoubtedly be healthy and could lead to a better and fairer system more in line with international standards. The ministry's excuse (and the excuse of review committee members) that there would be language problems is inadequate, at least for the ministry's larger grants, which are supposed to support "internationally recognized research". For these, assessment involving non-Japanese scientists would seem to be a prerequisite. □

Values poisoned by commerce

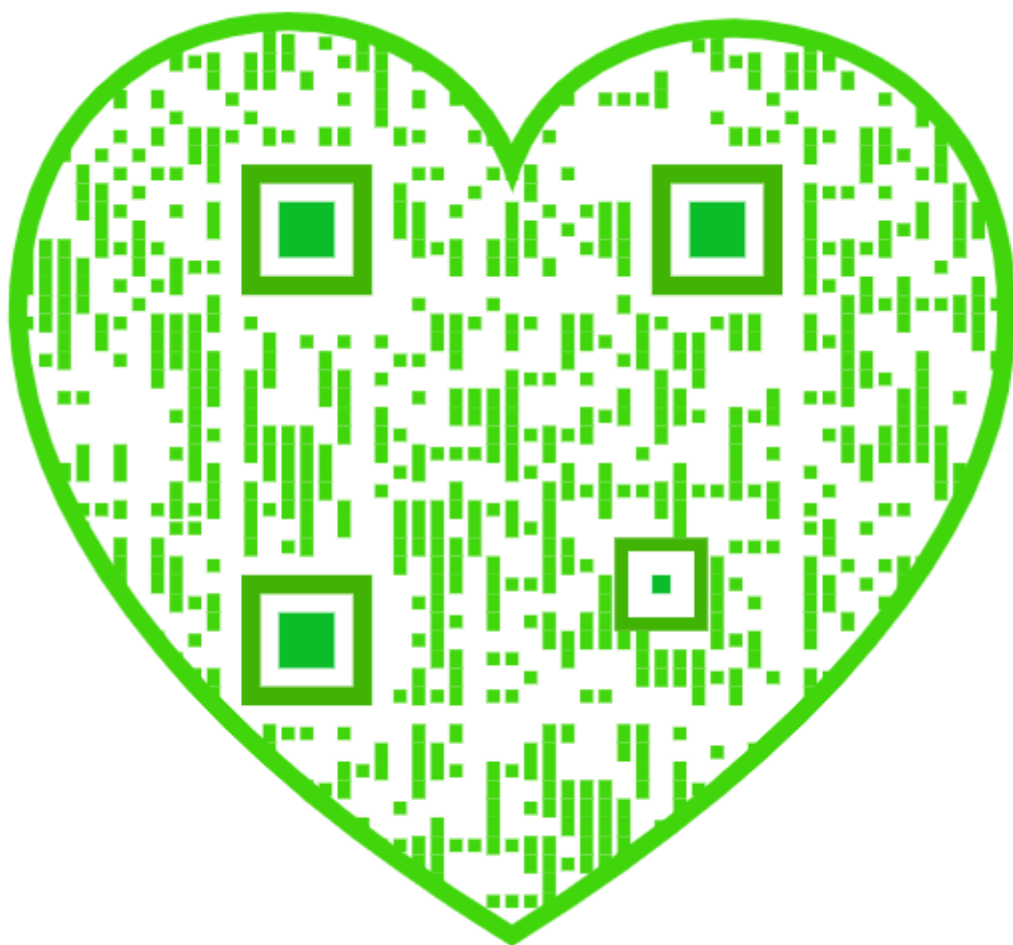
A survey and a 'leak' are symptoms of a corruption of traditional science.

OPENNESS and trust are traditional virtues of science, which its defenders and celebrators like to emphasize. In the end, reproducibility and falsifiability are the principal characteristics of the discipline, but in the practice of science these are intimately bound up with its cultural values. That the ideal is often undermined by personal ambition and commercial confidentiality is a misfortune that science has weathered over the centuries, whereas collisions with religious authorities and dictators have presented deeper threats. Another is emerging within molecular biology.

Today's fundamental elucidation of a protein interaction can be (almost literally) tomorrow's profit-earning test kit. Today's gene sequence, if not applicable tomorrow, may be assumed to be replete with vast profits achievable the day after. A survey (see page 574) shows that companies are acting on such assumptions, as their shareholders would no doubt insist they should. Rumours suggest that some individuals are doing the same (see page 574), as their colleagues would insist they should not. Although commerce is to be supported, there are signs of a culture being made sick by commercial interests to an extent that is unprecedented in science. The sociologists have a boom time ahead of them analysing the symptoms, but there is no sign of a cure, and the costs will have to be borne by everybody involved as the culture changes to one of legalistic precaution. Gallingly, those costs will serve to subsidize the relative few who make the profits. □

Washington move

The *Nature* editorial office in Washington DC will be at the following address from Monday 20 February:
968 National Press Building,
Washington DC 20045-1938
Telephone, fax and e-mail remain the same.



Small telescopes may be auctioned off under US 'privatization' plan

Washington. Most of the small telescopes at present available to all US astronomers would be closed or sold to private operators under a restructuring proposed for the National Optical Astronomy Observatories (NOAO). NOAO, which includes Kitt Peak National Observatory in Arizona and the Cerro Tololo Inter-American Observatory in Chile, hopes to compensate for the loss by building three new instruments for night-time observing. These would be a 2.4-metre telescope for each hemisphere, and a 4-metre telescope in Chile, built in cooperation with the University of North Carolina and with Brazil.

Another instrument for solar observations would be built at Kitt Peak and Sacramento Peak in New Mexico. The Association of Universities for Research in Astronomy (AURA), which runs NOAO, last month asked the National Science Foundation for \$21.6 million to cover the new construction work.

Faced with shrinking astronomy funding from the National Science Foundation, NOAO says it has no choice but to close or transfer eight of its smaller, older telescopes in Arizona and Chile. The only instruments that would remain in operation are the 4-metre Mayall and 3.5-metre WIYN telescopes on Kitt Peak and the 4-metre Blanco telescope (and perhaps a 1.5-metre infrared telescope) at Cerro Tololo. In addition to the three new telescopes to be completed by 2001, NOAO users would have access to the twin 8-metre Gemini telescopes, the first due to begin operating in 2000.

The proposals follow a recommendation by the National Research Council (NRC) last year that NOAO should place top priority on building advanced new telescopes, and keep the older telescopes open only if

additional money is available (see *Nature* 373, 179; 1995). But NSF's astronomy division has told NOAO to plan for level funding at best, and so the restructuring plan is designed to trim \$1.35 million a year from an annual operating budget of \$27 million.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Scanning the skies: Kitt Peak National Observatory would gain a telescope to compensate for closures.

Further savings might be made if more partners buy into the proposed new telescopes.

Sidney Wolff, the director of NOAO, admits that closing telescopes means fewer observers. But the addition of three new telescopes would mean that "a large fraction of current users still would be served," says John Salzer, an astronomer from Wesleyan University, Middletown, Connecticut, who chairs NOAO's user committee. The practice of 'queue scheduling', under which NOAO staff conduct observations rather than having astronomers visit the mountain themselves, would be one way of increasing the number of users.

But advanced technology and streamlined operations can go only so far. "Obviously, if you're going to close four telescopes and replace them with one, there's no way in the world you can get four times the science throughput," says Salzer.

New funds bring Magellan project into focus

Washington. The \$70-million Magellan project at Las Campanas in Chile is now officially a 'two-telescope' project, following a promise announced last week of funds from two more universities that will join the project's three existing backers.

The University of Michigan and the Massachusetts Institute of Technology (MIT) are joining the Carnegie Institution in Washington, the University of Arizona and Harvard University to make up an expanded five-member consortium. Telescope observation time will be proportional to an institution's investment in capital and operating costs. Carnegie, which owns the

present Las Campanas site, will be entitled to 50 per cent of observation time and will remain the project's 'lead partner'.

The other institutions will each be entitled to 10 per cent of the observing time, the remaining 10 per cent reserved for Chilean astronomers free of charge. A completion date for the telescopes has not been announced. But the primary mirror for Magellan 1 has been cast and is "ready for grinding and polishing" and scheduled for installation around mid-1998.

The telescopes will be situated at an elevation of 2,400 metres, far from any city lights and surrounded by stable air. □

NOAO observatories are used by nearly 1,000 researchers a year. Half of all US astronomers rely on them exclusively, as these are the only telescopes open to the entire community. Private observatories, although more numerous and in many cases more capable, are open only to astronomers whose institutions pay for the facility. Furthermore, says Salzer, "the majority of NOAO users are [users of] smaller telescopes".

The restructuring means that many of these astronomers will have to fend for themselves. One option is for smaller institutions to pool their money to buy time at larger observatories. Salzer's own university, Wesleyan, for example, has joined several others to buy time from Case Western Reserve University of Cleveland, Ohio, which owns one of the Kitt Peak telescopes.

Wolff expects that such consortia will eventually purchase the NOAO telescopes. Two previously closed telescopes on Kitt Peak have already been sold, and negotiations are under way for the sale of a third.

But Wolff warns that the "downside" to privatization is that "you do not have open competitive access for the community as a whole. We're moving to more and more restricted access to facilities, and that's a fundamental policy issue that the NSF really needs to consider."

The NSF is aware of the problem, says Hugh Van Horn, director of its astronomy division, and is considering the idea of allowing public access to private telescopes in exchange for NSF financing of advanced instrumentation — a key recommendation of last year's NRC report. "I regard this as one of the highest priority items" for NSF support of astronomy, says Van Horn. But this would apply only to the large telescopes now under construction, and is still far from sure, given NSF's funding uncertainties.

Even though the smaller NOAO telescopes are still very much in demand, Salzer agrees that building new, advanced instruments should have priority. "No matter how you cut it, the telescopes at Kitt Peak and [Cerro Tololo] are old, and they're not going to be viable telescopes for much longer." NOAO has found that it would be cheaper to build new telescopes than to upgrade the old ones to meet modern performance standards. In such circumstances, Salzer claims that the NOAO restructuring proposal — with its promise of three new telescopes — is probably the best option in an era of downsizing.

Tony Reichhardt

Republicans want 'junk science', says Gore

Baltimore. Vice President Al Gore this week sought to bring US science policy into the lengthy campaign for next November's presidential and congressional elections with a fiery denunciation of Republican science policy which, he said, had "the wisdom of a potted plant".

Gore said that Republicans were proposing "the sharpest cuts in non-defense research since World War II" and that "on question after question, they say that 'we don't know enough to act, and we don't want to know'". He added that recent statements from Republican congressmen "ought to send shivers up our spines" and suggested "a deeper disregard for science itself, a taste for junk science, and a belief that science is just another rhetorical tool to be twisted" for political ends.

Speaking at the annual meeting of the American Association for the Advancement of Science (AAAS) in Baltimore, Maryland, Gore sought to draw a clear line between Democrat and Republican visions for science. "At the end of this century we have a choice of two paths," he said. The Republican one was "a direct descendant of the society that insisted that the Earth was flat" in which research in key areas was neglected and "the houses of knowledge fell into disrepair". In contrast, the Democratic administration's path was "dotted with investments

that open the door to education for all of our people," he said.

Gore's speech to the AAAS was the first of three major science policy addresses he was planning this week. The others were to be made at a computer industry meeting in Crystal City, Virginia, and at the fiftieth birthday celebrations of the first US computer system in Philadelphia, Pennsylvania.



Gore: warned of 'deep disregard for science'.

The speeches come in what is effectively the opening week of the 1996 election campaign, with President Bill Clinton in Iowa campaigning against a still-to-be-selected Republican opponent, while congressmen return to their districts to prepare for elections in which the Democrats have an outside chance of regaining control of the House of Representatives.

One case that Gore cited as an example of Republican "disregard for science" was a comment made by Representative Tom DeLay (Republican, Texas), to the effect that the ban on CFCs was scientifically debatable and "the result of a media scare". Another was the memorable description

of global warming as "liberal clap-trap" by Representative Dana Rohrabacher (Republican, California), and a third example was a statement by an Oklahoma congressman that an outbreak of *Cryptosporidium* from a contaminated water supply was "helpful for finding those who are immune-compromised".

The vice president praised Jack Gibbons, Clinton's science adviser, as "the most influential science advisor the White House has seen for a generation", and, for the first time, sought to list the science policy achievements of the administration's first term. These included improved protection of intellectual property rights, increased funding for the major science agencies, more environmental research and better environmental regulation. The list excluded Clinton's technology programmes, which are now fighting for survival.

In the 1992 campaign, Gore had some success in associating the Clinton campaign with technology in general, and the 'information superhighway' in particular. This week's speeches suggest that Gore sees an opening for Clinton and himself to present themselves as the protectors of science against barbaric Republican cuts — a theme that will fit in with the broader education and environment planks of the Clinton presidential campaign.

Colin Macilwain

Health secretary backs AIDS office's budget authority

Washington. Donna Shalala, the US Secretary of Health and Human Services (HHS), insisted last week that, despite pressure from Republicans in Congress and a recent decision by the National Institutes of Health (NIH), the NIH's Office of AIDS Research (OAR) will not lose control of \$1.4 billion a year in research funding.

Addressing the annual meeting of the National Association of Science Writers in Baltimore, Maryland, Shalala said: "Make no mistake about it: the president and I are absolutely committed to maintaining a strong Office of AIDS Research — with its budget authority kept completely intact."

NIH director Harold Varmus echoed that commitment on Monday (12 February), but said there was no question of 1996 funding being routed through the OAR. "This year, I think it's done," he said. "But we will be making a very strong case for [routing money through the OAR in the] budget for [1997]". Varmus added that he was "strongly in support of the OAR and what [director] Bill Paul is doing".

Paul said that the decision to fund institutes directly in 1996 made little difference to the power of his office. "The great authority of the OAR lies in its responsibility to develop a comprehensive

plan for AIDS research and, secondly, to determine what sums of money the individual institutes will require to carry out the AIDS research activities that the OAR has approved," he said.

The OAR develops budget figures for AIDS spending by each NIH institute and centre early each year. The circumvention of the OAR means that the sums recommended for each institute, once appropriated by Congress, cannot be readjusted by the OAR, as they flow from the US Treasury to the institutes.

Various institute directors at NIH are known to resent the role of the OAR which, in 1996, is charged with distributing \$1.4 billion allocated to AIDS research, 11.8 per cent of the NIH's \$11.9-billion budget. But Shalala insisted that the 1993 law giving OAR authority to oversee all AIDS research at NIH would continue to be followed.

She said that the law should take precedence over a stop-gap spending measure passed in January funding NIH up to the end of the fiscal year. This has been interpreted by NIH officials, HHS lawyers and lawyers in the administration's Office of Management and Budget as directing that the OAR should be bypassed, and that research funds for 1996 should be

channelled directly to the NIH's 24 institutes and centres (see *Nature* 379, 475: 1996). The NIH says that the decision was taken in consultation with its lawyers.

Shalala claimed later that any scientists and administrators lobbying Congress to circumvent the OAR would, at least in theory, be putting their jobs at risk. "If I catch anyone, I'll fire them," she threatened. "But I won't catch them," she added.

On the same day (9 February), Representative John Porter (Republican, Illinois) insisted that the NIH's decision at the end of last month to channel AIDS money directly to its institutes was entirely "mechanical", and that it would have no impact on the political authority of the 33-member OAR.

But this interpretation is challenged by many AIDS activists and researchers, who say that the move would undermine the power of the OAR. "We don't want somebody [at the head of OAR] just saying 'it would be nice if you'd study this,'" says Gregg Gonsalves, policy director of the Treatment Action Group, an AIDS group based in New York. "We don't want [Paul] to be the king of AIDS research, [but we do] want him to be the prime minister," he said.

Meredith Wadman

No resuscitation for US technology office closed by Congress

Baltimore. Efforts to attract private finance to revive the former Office of Technology Assessment (OTA), which the Republican-dominated Congress voted to close last September, have come to nothing, officials told the annual meeting of the American Association for the Advancement of Science in Baltimore, Maryland, last weekend.

Vary Coates, president of the Institute for Technology Assessment (ITA), which has been trying to pick up where the OTA left off, said that that philanthropic foundations did not want to give Congress the impression they were prepared to pay for work that Congress had chosen to jettison. The search for funds has "been a real disappointment", Coates said, as the foundations "are not in the institute-building business". Nor were they in a mood "to pick up functions that Congress is [cutting] off", for fear of encouraging further cuts.

The OTA, set up in 1973 to advise Congress on matters ranging from patenting of human genes to nuclear waste disposal, was shut down by Republican congressmen who challenged both the value and objectivity of its work, and wanted to be seen to be cutting Congress's own costs. It had 200 highly skilled staff and a \$20 million annual budget.

The ITA is now seeking grant support for individual technology assessment projects. It has no full-time staff, but has a small Washington office supported by a start-up grant from the Garfield Foundation, a trust connected to the Philadelphia-based publishers of *The Scientist*, a weekly newspaper.

Roger Herdman, the last director of the OTA, said that the organization's work was not being done by any other congressional agency. The only prospect for continuing it lay in "privatizing" it, he said. David Guston of Rutgers University, New Jersey, said that a "plausible opportunity" existed for the ITA to occupy a niche, selling technology assessment to the 51 state governments.

Todd La Porte, OTA vice president, reported that five European bodies set up in OTA's mould — in Denmark, France, Germany, the Netherlands and the United Kingdom — are "progressing nicely", although each body is coming to reflect the political system in its own country.

The weakness of the French organization, for example, reflects the weakness of the French parliament, suggested La Porte. Organizations in Denmark and the Netherlands work directly to engage the interest of the general public, reflecting the participatory nature of democracy in these countries. "If there is less of a role for fact-based decision-making in the US Congress," he said, "the Holland-Denmark model might be worth taking a look at". **Colin Macilwain**

Cancer link to power cables 'exaggerated', say critics

London. The publishers of a peer-reviewed medical journal, and the Medical Research Council (MRC), have provoked controversy by helping to publicize a television programme that, some claim, exaggerates the results of a study of potential links between electromagnetic radiation and cancer.

Advance publicity for the programme, due to be broadcast on Wednesday night in Channel 4's series 'Dispatches', suggests that the researchers, whose work was carried out at the University of Bristol and supported by the MRC, have found such a link.

A notice issued to the press says that the team's results go "further than ever before" in showing how "electromagnetic fields surrounding both power lines and domestic mains leads to certain types of cancer". It

stand by its content as well as that of the television programme, and he denies that the MRC was "hyping the research".

Richard Horton, sales and marketing director for Taylor & Francis, also defends the press notice, pointing out that it had been circulated with the knowledge and consent of all concerned. "Some requested changes. Some changes were made," he says. "As publishers of the article and not the scientific expertise behind it, my advice is to see the programme."

Researchers have for some time been investigating claims of a possible link between power lines and cancer. So far, however, the evidence remains patchy. Epidemiological studies show an association between exposure to electromagnetic fields and cancer. But studies at the cellular level have failed to establish that such fields are the cause of the observed disease.

James Metcalfe, of the department of biochemistry at the University of Cambridge, was a participant in a three-year research project that tried unsuccessfully to replicate a much-cited study that claimed exposure to electromagnetic fields increases the risk of leukaemia (see *Nature* 375, 22; 1995). Although Metcalfe says the Henshaw paper is not wrong to speculate about a potential link between radon decay products, electric fields and cancer, he points out that the paper emphasizes that a cause has yet to be established, and describes the publicity for the paper as "oversimplified".

But Michael O'Riordan, a spokesman for the National Radiological Protection Board, says the theory being put forward by Henshaw and colleagues — namely that electric fields can enhance the effects of inhaled radon — is "implausible". O'Riordan says that the effect of electric fields on radon decay products, which decay from uranium in the ground and rise into the atmosphere as part of soil gas, has been known since 1901. He also acknowledges that a link has been established between radon decay products, known as 'daughters', and lung cancer.

But O'Riordan says that Henshaw's proposal conflicts with established scientific literature. Electric fields, he says, cause radon decay products to stick to surfaces. "The consequence is that fewer radon daughters remain in the air to be breathed by occupants, so exposure is lowered."

At the beginning of this week, Henshaw and the Bristol team said that they had agreed with the MRC not to comment on reports about the programme's content until its advance screening on Tuesday (13 February), despite widespread reports in newspapers based on the earlier press notice.

Ehsan Masood

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Christopher Priest/SPL

Over the top? 'No persuasive evidence' of a link between electric fields and cancer.

adds: "evidence of how these fields could cause cancer has previously baffled scientists — until now". One subsequent headline in the *Times* read: "Scientists link power lines to cancer".

But the research, published on the same day in the *International Journal of Radiation Biology*, explicitly denies that it has shown any such link. In their paper, "Enhanced deposition of radon daughter nuclei in the vicinity of power frequency magnetic fields", Denis Henshaw and colleagues speculate about a possible link between certain types of cancers and the inhalation of radon decay products that are influenced by electric fields. But they state in their abstract that "at present, there still remains no persuasive biological evidence that power frequency electromagnetic fields can influence any of the accepted stages in carcinogenesis".

Paul Fawcett, a spokesman for the MRC, says that the press notice, circulated by Taylor & Francis, publishers of the journal, was intended primarily as an invitation to an advance screening of the documentary. He adds that both Henshaw and MRC were slightly "annoyed" about the wording, but

INTAS plans improved grant systems after claims of 'patronage'

Brussels. INTAS, the European organization that supports research projects in New Independent States (NIS) of the former Soviet Union, is to streamline its procedures for handling grant applications in order to meet criticism that it is too bureaucratic, too slow and too secretive.

Pierre Venet, the general secretary, told the general assembly of INTAS in Brussels last week that mechanisms are being set up to ensure that money is transferred to the NIS scientists safely and quickly, and that a full list of approved projects will soon be made available.

He also promised that procedural guidelines for the Council of Scientists, the body responsible for recommending which projects are funded, will be provided before the next call for proposals at the end of this year and that these will stipulate that council members must not be involved in INTAS projects. Concern had been expressed at the assembly about rumours that some members of the council had either received or been personally linked to a large number of approved projects.

Niceas Schamp, the chairman of the council and an organic chemist from Gent in Belgium, denies that this has been the case. But a list of the number of projects with which council members are directly associated indicates that two members have links with 53 and 37 projects respectively.

Both INTAS and the European Commission (EC), which chairs the assembly and provides most of the funding for the programme, deny that patronage by some council members has skewed the allocation of grants — and they are concerned by rumours that this has happened.

In addition, they point out, for example, that those individuals concerned are world renowned scientists, and that it was therefore not surprising that they were appointed to the council and were successful in winning grants.

But Annalise Eggimann, Swiss representative to the INTAS assembly, says that both INTAS and the commission should indicate that the critics, who include individuals inside both the commission and INTAS, are being taken seriously by providing statistics on the involvement of council members in projects.

Concern was also reflected in the decision taken by the INTAS assembly to select new members of the committee before the next call for proposals.

Alison Abbott

Nuclear watchdog tightens incident reporting rules

Washington. Suspicious incidents involving radioactive materials in research laboratories would have to be reported at once to the Nuclear Regulatory Commission (NRC) under a new rule proposed by the nuclear watchdog agency.

The NRC wants to be informed within 24 hours of "any intentional, or suspected intentional, diversion of radioactive material from its intended or authorized use". Unlike existing regulations, the new rule is



intended to 'catch' incidents that cause very small radiation exposures — or even the hypothetical risk of such an exposure.

The agency said it was proposing the rule because of two incidents last year — one at the National Institutes of Health (NIH) in Bethesda, Maryland, the other at the Massachusetts Institute of Technology (MIT) — in both of which researchers from China ingested sizeable amounts of phosphorous-32, a common radioactive tracer. No satisfactory explanation has yet been found for

either of the incidents, which are still under criminal investigation by the Federal Bureau of Investigations (FBI) and the NRC.

The NRC was informed immediately of the NIH incident. But it learned of the second incident from MIT only six weeks after the event, when the story was about to break in the media (see *Nature* 377, 568; 1995). In each case, the dose of radiation received by the victim fell slightly below 600 rems, the level at which the laboratory was obliged to alert the NRC under old rules.

Joe Gilliland, a spokesman for the NRC, said last week that the agency "might have been able to get on top of the situation a little earlier" if it had been informed of the MIT incident at once. "A lesson we learned there was that we needed to sharpen up our reporting practices," he said. NIH did report its incident the day after it happened, Gilliland added; but "there was a question of whether it was required to do so" under the old rules.

The proposed rule was published in the Federal Register on 31 January and interested parties have until 1 March to comment on it. The two apparently unrelated incidents have already resulted in sharp crackdowns by the NRC at both NIH and MIT. Many researchers are irritated by constraints — now strictly enforced — on access to radioactive tracers and the rooms in which they are used, and on eating and drinking in these rooms.

Last December, a report from the Institute of Medicine said the NRC should be stripped of responsibility for regulating radioactive materials in medicine and in biomedical laboratories. The report argued that the NRC did not understand how medicine worked.

Colin Macilwain

Ministry 'knew high risk' of AIDS transmission

Tokyo. A three-day search in Japan's Ministry of Health and Welfare has proved what the government has previously refused to admit, namely that in 1983, the ministry was already aware of a strong possibility that the infectious agent causing AIDS is transmitted through blood and bodily fluids.

The search was ordered by Naoto Kan, Japan's minister of health, soon after he took office last month in the new cabinet of Prime Minister Ryutaro Hashimoto. On 25 January, 72 people, including HIV-infected haemophiliacs, filed a law-suit with the Tokyo District Public Prosecutor's office accusing a former ministry official, Atsuki Gunji, of perjury in earlier court hearings, when he said that he had been unaware in 1983 of the likely similarity of the infection routes of the AIDS-causing

agent and hepatitis B virus (see *Nature* 379, 388; 1996).

The group claimed to have a copy of a document submitted to the ministry's AIDS study group in July 1983 that proved their accusation. But the ministry said at first it could not confirm this paper's authenticity as it had lost all such documents. Records of the AIDS study group meetings have now been found in the ministry library. They reportedly state that the causative agent of AIDS is "highly likely transmitted via blood and other bodily fluids".

Announcing the find last week, Kan said that the documents support claims that the ministry was aware of the infection route, and therefore bears some responsibility for the infection of thousands of haemophiliacs in Japan with HIV.

David Swinbanks

Japan sharpens selection of grant winners

Tokyo. The Japanese government has announced changes in the system for awarding competitive research grants to university researchers that will, it hopes, increase the chances of success of highly rated grant applications, and also ensure that the selection procedure becomes more transparent.

Under the reforms, approved on 1 February by the Science Council of the Ministry of Education, Science, Sports and Culture (Monbusho), which funds the bulk of university research, all grant applications will be given absolute rather than relative evaluative scores, while the names of members of review committees — currently kept secret — will be revealed.

But, while welcoming the reforms, critics say that much remains to be done if Japan's procedures for awarding grants are to become comparable to those followed in other countries. In particular, there are demands for more feedback on failed grant applications and a greater role for foreign scientists in making grant decisions.

The budget for Monbusho's grant programme has grown dramatically in recent years. In the fiscal year 1996, from April, it will top ¥100 billion (about US\$1 billion) for the first time, almost twice the level of only five years ago (see figure, above right).

But the grant system has long been criticized by university researchers — even those who have won large grants — for its lack of both transparency and feedback (see, for example, *Nature* 340, 494; 1989). Both defects have fanned suspicions that researchers with personal connections to powerful members of review committees

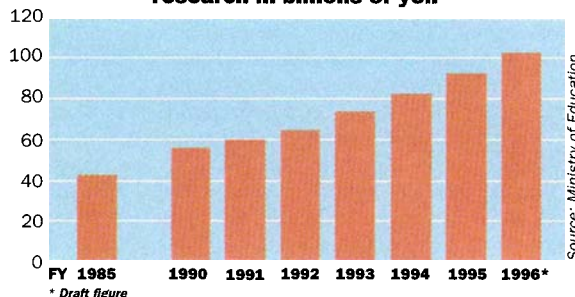
have a better chance of success than others.

The reforms now adopted by the Science Council will allow referees to award high marks to all deserving applications, rather than merely a pre-defined percentage. At present, a grade between one and five is given to each application and a set portion (typically 20 to 25 per cent) with high grades are given grants. But the grading system is not consistent across fields.

One consequence is that in fields in which Japanese research is weak, projects of a relatively low standard may be funded while high-quality applications in fields with a greater overall level of achievement are rejected. The new rules will, the government hopes, alter this situation. Other changes include a reduction in the number of grant categories, and the decision to publish the names of all grant review committee members once they have completed their two-year terms of duty.

Many university researchers have welcomed the reforms, even if they are relatively modest. Mitsuhiro Yanagida of the biophysics department of Kyoto University, for example, says that the plan to publish the names of the reviewers of grant applications is a significant improvement. Takeyuki Wakabayashi, of the faculty of science at Tokyo University, points out that the names of referees, while officially secret, are in fact well known to senior scientists in Tokyo, but

Japanese government's budget for scientific research in billions of yen



often not to researchers outside the capital.

Monbusho officials, as well as academics who have served on the review committees, insist that all applicants are treated equally. But many still feel that reviewers may favour those they know personally. Wakabayashi hopes that the new grading system will allow reviewers to award high marks to all worthy applications and a higher proportion of good applications will receive funding.

Members of the review committees, many of whom are from Tokyo University, dismiss critics of the grant screening process as "failed applicants". But in fact some of the most severe criticisms of Monbusho's system have come from scientists who have won large grants (see, for example, *Nature* 343, 111; 1990).

Indeed, many would like the ministry to go further and enact other reforms, such as the disclosure of referees' comments on grant applications. Even under the new system, for example, applicants will receive no feedback on the success or failure of their applications.

A Monbusho spokesman says that the ministry is discussing whether to provide feedback to a hundred or so applicants who apply for especially large grants. But he says that the large number of applications — about 90,000 per year — makes it impractical to provide feedback to all applicants.

Inclusion of non-Japanese scientists in the grant review process is also desirable, says Yanagida. But the language barrier remains a problem. Although leading Japanese scientists are often asked to assess grant applications in foreign countries, it is difficult for non-Japanese to be consulted on the selection of domestic grants because applications to Monbusho are made in Japanese.

Government funding for science has grown rapidly in recent years, making the need for improved evaluation of grant applications even more pressing, says Akito Arima, a former president of Tokyo University and current president of the Institute for Physical and Chemical Research (RIKEN). Arima, who describes himself as "one person who has insisted" on reform, says he is very happy with the announced changes, and he expects further reforms in the near future.

Stephen Barker

Pacific gulf limits US-Japanese projects

Washington. Fundamental differences in outlook continue to inhibit effective scientific collaboration between the world's two largest economies, leading scientists from Japan and the United States told a forum in Washington last week on science in Japan.

The forum was told that there remains a huge gulf in the number of researchers and students exchanged between the two countries. Japan has 12,224 researchers in the United States, while only 3,131 US researchers work in Japan, according to Hitoshi Osaki, director of the Japan Society for the Promotion of Science (JSPS), which organized the meeting.

But scientists from the two countries have different views on solving the problem. Osaki suggested that "the most important thing is to have attractive research centres", and that "the Japanese government is now beginning to

realize the fact". But US scientists said that language problems, other cultural factors and high living costs prevent more US researchers from working in Japan.

Minoru Oda, one of Japan's leading astronomers and president of the Tokyo University of Information Science, said that although individual contacts have to form the basis of collaborations, the rules imposed by US funding agencies inhibit this process. In contrast to joint projects with British scientists, for example, US science funding agencies insist on vetting every collaboration at the outset to ensure that it complies with umbrella agreements.

US delegates appeared envious of budget increase that Japanese science will receive this year, but criticized the inflexibility in the Japanese research system, especially as it affects young scientists and women.

Colin Macilwain

'Leak' rumours fuel debate on gene patent

London. Two research groups potentially engaged in a priority dispute over the discovery of the second hereditary breast cancer gene, *BRCA2*, have entered negotiations over the terms under which each may let the other look at data used in preparing patent applications that were filed only a few days apart.

The negotiations follow the announcement on 20 December by Myriad Corporation of Salt Lake City that it had filed a patent application for all the diagnostic and therapeutic uses of the gene, the day before its details were first published by researchers at the Institute of Cancer Research (ICR) in

London (see *Nature* 378, 789; 1995).

The British group has also applied for a patent. In addition to indicating whether the patent claims are in conflict, the British scientists may be hoping that access to Myriad's patent data will throw some light on the truth or otherwise of rumours that the US team was assisted in its search for the full *BRCA2* sequence by a 'leak' of critical information from one of the teams with which they were collaborating.

Such speculation is based partly on the speed with which Myriad, believed only days previously to have been investigating a number of candidate genes, said it had success-

fully located *BRCA2*, using the full sequence made publicly available over the Internet by the Sanger Centre in Cambridge, England and the University of Washington in St Louis (see *Nature* 378, 425; 1995).

Gordon McVie, director of the Cancer Research Campaign which provided much of the funding for the ICR work, says that Myriad's announcement was no surprise, as it had been working hard to find the gene. But he acknowledges that he is aware of the rumour of a leak, said by some to have centred on the crucial information that *BRCA2*, like its companion *BRCA1*, contains one unusually large exon.

Scientists from Myriad, whose own data are due to be published in next month's *Nature Genetics* — and who had already been among the favourites to find *BRCA2* following their earlier success in locating *BRCA1* — are refusing to comment on the company's announcement that it had discovered the gene and applied for a patent on it.

But Chris Taylor, a spokesman for the company, says that he "strongly disagrees" with the claim that Myriad researchers had made use of data derived from the ICR team in their discovery. Taylor points out that, while Myriad acknowledges that the ICR team, headed by Mike Stratton, was the first to publish what it calls the "partial sequence" of the gene, it was Myriad which first made the whole sequence available through GenBank.

While legal representatives of the British and US scientists investigate the degree of overlap between their patent claims — and thus weigh up whether to challenge each others' claims in court, or to seek some mutual accommodation — others have been commenting on the use by both teams of the Sanger/Washington University data.

Some are concerned, for example, that Myriad's substantial investment in gene sequencing and other analytical equipment has placed it in a much stronger position to make use of such 'public' data compared to smaller, less well-resourced groups.

"I am worried that anyone who asks one of these 'public' data banks in future to sequence a human gene may, as a result of such a policy, immediately find the information produced being patented by a large pharmaceutical or gene sequencing company," says one scientist, adding that in future he intends to keep all his own sequence information private until it is protected.

But David Bentley of the Sanger Centre defends the decision to go public with the full sequence within which the *BRCA2* gene was later found by both the ICR and Myriad teams. "If we hide this data, we will not be allowing the maximum number of people to use it as they can, and we will be deciding who should have access to it," he says.

David Dickson

Commercial interests delay publication

Washington. US companies that sponsor university research projects often require the scientists involved to keep results secret longer than patent procedures require, according to a survey of 210 companies published last week in the *New England Journal of Medicine* (*NEJM* 334, 377; 1996). Data on gene products, sequences and location are among the strategically valuable information withheld.

The respondents included major pharmaceutical, agricultural, chemical, biotechnology and other companies. Led by David Blumenthal, chief of health policy research at Massachusetts General Hospital, the study found that 59 per cent of the companies funded life sciences research at US universities. Their total spending of about \$1.5 billion in 1994 represented about 11.7 per cent of all research and development funding at the universities.

More than half (58 per cent) of the companies said they typically require university researchers to refrain from publishing results for at least six months in order to file patent applications. By comparison, the National Institutes of Health (NIH) has said that 30 to 60 days was a reasonable time for its researchers to be required to keep results confidential while patent applications are filed.

A slightly lower proportion (56 per cent) admitted that their university-supported research often or sometimes generated information that was kept secret beyond the time required to file a patent. Such information included experimental methods (listed by 56 per cent of respondents), plans for future experiments (53 per cent), gene products (27 per cent), gene sequences (25 per cent) and gene location (12 per cent).

The study has met with mixed reactions from academic researchers.

Blumenthal says that secrecy should be "minimized", but calls it "a more or less inevitable concomitant of industry-funded research".

Some accept that corporate confidentiality requirements are reasonable, given a company's need to remain competitive. "In the real world, there are people out there trying to pick off somebody else's discovery, and that's where the problem is," says Ronald Breslow, president of the American Chemical Society who, as an organic chemist at Columbia University, has collaborated with industry.

Breslow says, for example, that when he discovered a novel anti-cancer effect of a particular compound in tissue culture, it was entirely reasonable for his industry sponsor to ask him to keep the result confidential until it had been able to test the compound for toxicity and effectiveness in animals.

But others claim that commercial secrecy requirements have increased the frequency with which researchers attending scientific meetings refuse to disclose all their information. Sheldon Krinsky, a science policy specialist at Tufts University in Medford, Massachusetts, who has studied the relationship of industry to academic science, says that the new data suggest that companies that support scientists are asking for "a lot more" than simply the protection of information to gain patents.

Others add that companies can require scientists to keep silent for as long as a decade. In an accompanying commentary to the *NEJM* article, Steven Rosenberg, a researcher at the National Cancer Institute, reported that one company had denied a request for a reagent after he had refused to agree not to disclose any data or results for ten years.

Meredith Wadman

University renames faculty after row over eugenics advocate

Paris. The board of the University of Claude Bernard in Lyons, France, has voted to remove from its faculty of medicine the name of Alexis Carrel, winner of the Nobel prize for physiology or medicine in 1912, whose reputation has recently come under fire for his support of eugenics in the years before the Second World War, and for his alleged links with the collaborationist Vichy regime (see *Nature* 378, 122; 1995).

But as one controversy abates, a related dispute has been opened up by the announcement by a historian at the Centre Nationale de la Recherche Scientifique that he intends to sue Presses Universitaires de France (PUF), one of France's leading academic publishers, for refusing to publish an article that it commissioned from him about Carrel.

Forty-six of the 59 members of the board of the Lyons university voted in favour of removing Carrel's name, which the university hopes will end the bitter controversy between the supporters and opposers of such an act. The university intends to rename the faculty RTH Laennec, after the less controversial René Théophile Hyacinthe Laennec (1781–1826), who invented the stethoscope and pioneered research on lung and liver disorders.

The complaint against PUF has been made by Alain Drouard, an expert on Carrel who has long argued that the biologist's eugenic views have been taken out of their historical context, in particular the widespread support for eugenic ideas among scientists in the 1930s.

Drouard's article was to have appeared in the three-volume *Dictionnaire du Darwinisme et de L'Évolution*, published last month. But Patrick Tort, the encyclopaedia's editor and an outspoken critic of Carrel, replaced Drouard's article with one of his own.

Tort claims he was merely exercising his right as an editor, and says he rejected the article because he felt its conclusions constituted "unacceptable" revisionism. But Drouard, who argues that PUF has broken a contract made with him in 1991, accuses Tort of "ideological censorship" as well as of infringing academic liberty and free speech.

Ironically, while Tort now argues that Drouard's paper was of "bad quality", he wrote to Drouard in 1991 saying that although the article was somewhat "kind" to Carrel, it was nonetheless "excellent". **D.B.**

French citations forge ahead but Europe still loses ground

Paris. France's share of papers appearing in international scientific journals increased by 14 per cent over the period 1982 to 1993, a rise that appears to reflect both increased research spending under the late President François Mitterrand and the increasing number of French scientists publishing in English. In contrast, Germany's share during this period decreased by 4 per cent, and that of the United Kingdom by 7 per cent.

But in overall figures, French publications still lag behind those of its two main European competitors, according to the 500-page third report of the Paris-based Observatoire des Science et des Techniques (OST). In 1993, for example, France was responsible for 4.9 per cent of the total number of international scientific papers, compared with 6.2 per cent for Germany and 8.5 per cent for the United Kingdom.

Furthermore, these figures mask much greater swings within individual disciplines. For example, although the United Kingdom maintains a strong position in biomedicine, accounting for 12.4 per cent of all international publications, its share of papers in applied biology and ecology fell by 22 per cent, and those in mathematics by 21 per cent. In contrast, France's share of mathematics papers increased by 59 per cent, from 4.2 to 6.7 per cent of the total.

Within Europe, the greatest gains have been registered by those countries that started from a lower base. For example, both Spain and Portugal more than doubled their share of total international scientific publications by member states of the European Union, to 5.6 per cent and 0.4 per cent respectively. Italy increased its share from 7.6 to 9.3 per cent.

One explanation for France's growing share of international scientific publications is the rejuvenation of French science under the government of Mitterrand, when public spending on research increased from 1.97 per cent of gross national product to 2.4 per cent at the end of the 1980s. (OST cautions, however, that funding for French science has "stagnated" since 1990.)

At the same time, France's increased share may also partly reflect a trend to publish more in English, in particular in mathematics. In 1982, 68.1 per cent of articles published by French scientists in journals included in the *Science Citation Index*, published by the Institute for Scientific Information in Philadelphia were in English; this figure has now risen to 81 per cent.

Furthermore, only 73.3 per cent of articles published by French scientists in French journals are now in French, compared with 84.5 per cent in 1982. They may have good reason to prefer publishing in English;

according to OST, French scientific journals that are published either partially or totally in English have an impact factor 3.5 times higher than those published in French alone.

In global terms, the OST report shows that the European Union (EU)'s share of the world's international scientific publications, based on a variety of sources that includes in particular articles cited in the *Science Citation Index*, increased by 7 per cent, to 31.4 per cent of the total, between 1982 and 1993. Over the same period, the US share of the total fell by 4 per cent (to 35.5 per cent), while Japan increased its share by 19 per cent (to 8 per cent).

Perhaps the most striking changes over this same period are the halving of the former Soviet Union's share, from 8.4 to 4.8 per cent of the total, and the quadrupling of the shares of both China (0.3 to 1.2 per cent) and the emerging economies of South-East Asia (0.3 to 1.4 per cent).

The report also confirms that Europe is losing ground to the United States and Japan in key areas of technology, at least measured in terms of patents. The strongest growth in technology has occurred in Japan and South-East Asia. But the United States still accounts for half the world's 'technological activity' — calculated by OST partly on the basis of the levels of public and private financing of research — in pharmaceuticals, computing and biotechnology.

In contrast, and using the same yardstick, Europe has generally fallen behind in key technologies, in particular in computing and telecommunications, although it has made progress in areas such as aerospace, pharmaceuticals and instrumentation.

Within Europe, Germany accounts for 37.6 per cent of the total of European patents taken out by EU countries in key areas of technology. France holds 23.6 per cent, and the United Kingdom 15.7 per cent. France leads in computing with a 29.7 per cent share. Moreover, the share of European patents taken out by EU countries in all areas has fallen by 9 per cent to 49.5 per cent over the decade. In contrast, the US share increased by 6 per cent (to 28.1 per cent), while that of Japan jumped by 22 per cent.

Within Europe, France's share increased from 17.3 to 18.1 per cent between 1987 and 1993. Over the same period, Germany's share remained stable at around 43 per cent, while the UK share fell by 17 per cent to 12.4 per cent. In terms of US patents taken out by EU countries, France's share increased from 15.8 to 17.7 per cent, whereas Germany's share fell by 10 per cent to 42.4 per cent, and the UK's share by 24 per cent to 14.3 per cent.

Declan Butler

British government acts to protect Internet providers against libel

London. The British government has introduced a bill into the House of Lords designed to protect 'innocent' Internet-providers from being sued for defamatory material that passes through or is held on bulletin boards. The new procedures, introduced by Lord Mackay, the Lord Chancellor, includes a defence of 'innocent dissemination'. The bill would also increase libel protection for the printing industry, as well as for shops selling newspapers and magazines that unwittingly help to disseminate defamatory material.

Legislation was first suggested seven months ago, when Britain's first Internet libel case was settled out of court (see *Nature* 375, 525; 1995). Laurence Godfrey, a physics lecturer, had secured a payment from another physicist for posting allegedly libellous remarks on the news group, Usenet.

Godfrey had said he would also sue several Internet providers who had allowed the remarks to remain on bulletin boards despite requests for their removal. Two providers in the United States apparently refused as they regard interference on the Internet as a violation of freedom of speech. □

Lenses endow neuroscience at MIT

Boston. Menicon Co. Ltd, a Japanese manufacturer of contact lenses in Japan, announced last week that it is to endow a professorship in neuroscience at the Massachusetts Institute of Technology (MIT).

A \$2.5-million grant from the company will be used to recruit a professor to the MIT Center for Learning and Memory that is headed by the Nobel laureate, Susumu Tonegawa. Kyoichi Tanaka, president of Menicon, has described learning and memory as "a very exciting new field in brain and cognitive sciences", adding his belief that "Tonegawa's centre will make significant discoveries that will be of benefit to people around the world. This is why

Menicon decided to take this unusual initiative."

The announcement has been welcomed by Emilio Bizzi, head of the department of brain and cognitive sciences, which provides three faculty positions for the centre for learning and memory. The centre has temporary space in the department of brain and cognitive sciences, and in a couple of years MIT will provide the centre with its own accommodation. □

Trilateral marine missions agreed

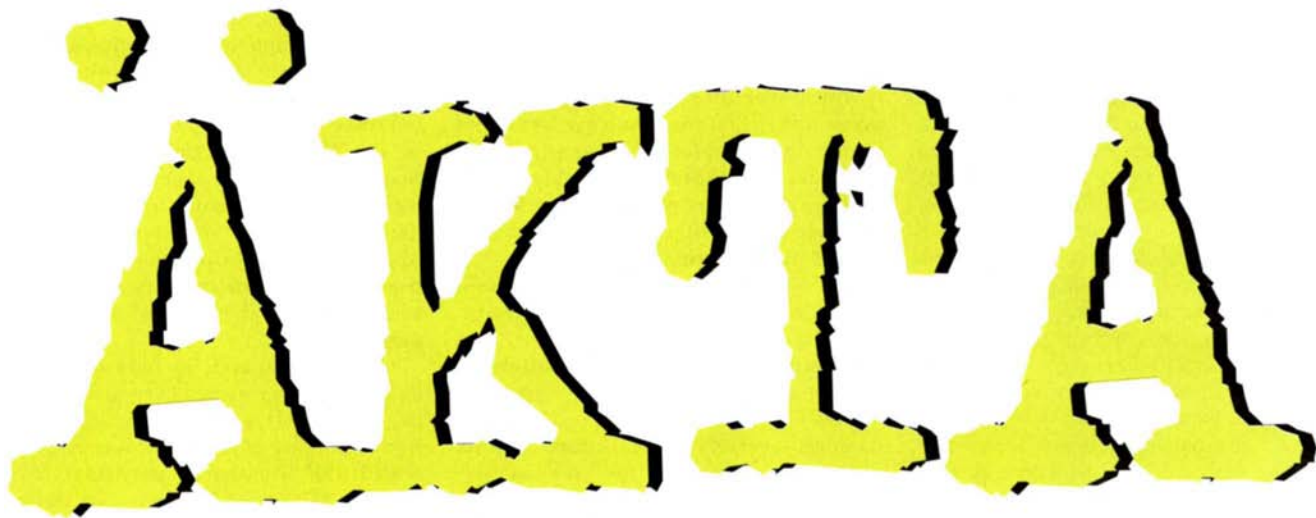
Munich. Germany, France and Britain this week signed an agreement to combine their marine research facilities, including 15 research vessels of various sizes, in order to contribute more efficiently to the expensive business of deep-sea research. This will mean that German, French and British scientists may join any of the ships' missions, all of which will be advertised in the three countries, and share each vessel's specialized equipment. □

Calgene survives patent challenge

San Francisco. Calgene Inc., the biotechnology company based in Davis, California, has survived a patent challenge to its best known product, the genetically engineered 'Flavr Savr' tomato. But, after disappointing trial results, production of the Flavr Savr has been reduced, and the company has postponed commercialization while it evaluates new transgenic tomato varieties.

A judge in Delaware has ruled that the company has not infringed three patents owned by Enzo Biochem Inc. of New York on antisense, the technology used to extend the Calgene tomato's shelf-life. The judge ruled that Enzo's patents do not enable others to make use of the claimed invention, and were therefore invalid; in contrast, he said, Calgene's patent was valid.

Last week, however, the company said in a quarterly report released to investors that the Flavr Savr did not have acceptable yield or disease resistance traits, and that its quality also needed improvement. Calgene officials claims that such efforts will be



Cost-cutting at CERN?

SIR — As a British physicist who spends about one third of his time working at CERN, I am in a position to appreciate the difficult situation that has developed, and I should like to comment on your recent News story (*Nature* 379, 286; 1996).

First, CERN's new project, the Large Hadron Collider (LHC), was unanimously approved by the CERN member states in December 1994, on condition that the budget be frozen during 1995–97 and thereafter indexed 1 per cent per year assuming 2 per cent inflation. This approval was given only after CERN had found ways to make substantial cuts in running costs, and reluctantly proposed to end excellent programmes (for example LEAR, where antihydrogen was recently discovered) in order to make resources available for the LHC, which has an even higher priority.

Second, because CERN's budget is determined in Swiss francs, the dramatic fall in the value of the pound relative to the Swiss franc has caused a serious problem for the United Kingdom. However, the strength of the Swiss franc has not created a large wind-fall gain for CERN because 75 per cent of CERN's expenditure is in Swiss francs (on salaries, utilities and so on), while the remainder is mainly in currencies (especially the French franc and Deutschmark) that have held up better than the pound relative to the Swiss franc.

Third, your reference to "a new salary award for CERN staff" gives a misleading impression and, incidentally, your statement that the United Kingdom voted against this year's budget is wrong: the United Kingdom voted against any indexation of salaries, not against the budget as such. In fact, CERN salaries were indexed 1.3 per cent inside a constant personnel budget, but 0.6 per cent was accounted for by increased contributions to the pension fund, leaving only a 0.7 per cent increase in take-home pay, compared with a 2.1 per cent rise in the Geneva cost-of-living index. It should also be noted that CERN salaries have been under-indexed during each of the previous four years. CERN should of course be encouraged to continue the significant gains in "productivity" that have been achieved over the past few years.

None of this is to deny that the rise of the Swiss franc has created a very serious problem for the United Kingdom (and other CERN member states), exacerbated by the rise in UK gross domestic product, which has increased the UK share of the CERN budget. Unfortunately, the savings resulting from continued salary restraint and cost cutting will not be on the scale of the United Kingdom's problem, a 27 per cent increase in the cost in pounds over two years.

As chairman of the Advisory Committee

of CERN Users, which represents the views of the 6,500 scientists who visit CERN to work on the experimental facilities, I am also aware of the complexity of the issues involved in deciding the appropriate salary levels, and that simply focusing upon one element of the salary settlement misrepresents the true situation.

K. J. Peach

Particle Physics Experiments Group,
Department of Physics and Astronomy,
University of Edinburgh,
Mayfield Road, Edinburgh EH9 3JZ, UK

Warped view

SIR — Cecil Fox's appeal (*Nature* 379, 292; 1996) that we should eschew all "fantasies" to abolish the public mistrust of science shows why the public are justified in their mistrust of at least some scientists.

John Maddox (*Nature* 378, 435–437; 1995), gave his first consideration to "exaggerated claims", but only to technology, not to ideology. Perhaps in his well-tempered valediction he overlooked the ability of some scientists to be blind to the boundary of their own expertise. To over-extend hard-wrought knowledge into a universal and singular interpretation of truth and value begs to be described as puritanical.

First on Fox's hit-list of heresies is the "trivial" television series 'Star Trek' which, he asserts with certainty, "would fail to appeal to a public forearmed with a belief in the tidiness of thermodynamics". Just how does Fox categorize all those thermodynamically aware Trekkies, some of whom are, doubtless, among your readers?

This school's current outline for its Year 7 (11–12 years old) course: 'Earth in Space' reads, under "know the extent of manned and unmanned space exploration":

"The film *Apollo 13* makes a useful topical departure point at present. It is important to point out that the Apollo missions were remarkable because they were real; also point out that 'Star Trek' may have good story lines but is nonsense science."

Of course, the likes of Cecil Fox would not need to have the scientific aspects pointed out but they should recognize that a good comprehensive school like this one should be able to cater for those who show exceptional ability in only one part of the curriculum. He does not have to appreciate the quality of the treatment in 'Star Trek' of aspects of humanity and metaphor, but he should not assume that everyone else who is scientifically literate would find good creative writing and acting "trivial".

Roger Macy

Mascalls School,
Paddock Wood, Kent TN12 6LT, UK

Organ transplants

SIR — Two recent letters about xenotransplantation^{1,2} require a response.

Mohacs *et al.*¹ stated that opposition to xenotransplantation has been expressed by the general community, and supported this statement by referring to a study carried out in the United States³. Far from demonstrating opposition to xenotransplantation, that survey indicates substantial support. Of 6,127 people surveyed, 85 per cent expressed general support for organ donation, and 79 per cent said they would accept an organ transplant if they needed one. The percentage approving the transplantation of animal organs when human organs were not available was 50 per cent, and 51 per cent would accept an organ transplant from an animal if a suitable human organ was not available. It should be remembered that these were the opinions of healthy men and women, who presumably doubted that they would ever need an organ transplant. My own opinion, based on 18 years in clinical heart transplantation, is that, if the survey had been carried out among patients awaiting a donor organ, the percentage who would have accepted a xenograft would have been very substantially higher.

Mohacs *et al.* also refer to their own study carried out among acute care nurses in Australia, who were clearly opposed to the use of xenografts. It would be interesting to learn the views of these nurses on the use of pig heart valves and pig insulin in humans, two common forms of therapy that have been used for many years.

In the second letter, Robin Holliday² drew attention to the fact that the life-span of domestic pigs may be relatively short, certainly not more than 20 years, and suggested that the success of xenotransplantation "might be completely foiled by the fairly rapid ageing of the transplanted heart". But one of the great advantages of xenografting would be that, if the organ fails (as unfortunately still occurs after allografting), a new organ is immediately and readily available. This remains the case with surgically implanted pig heart valves in humans which, despite their advantages, require replacement at intervals. Despite the inconvenience and small risk of undergoing sequential surgeries, these valves have saved many lives.

It is perhaps ironic that both of these rather 'negative' letters emanate from Australia, a country where much excellent research is being carried out in the field of xenotransplantation.

David K. C. Cooper

6700 NW Grand Boulevard,
Oklahoma City, Oklahoma 73116, USA

1. Mohacs, P. J. *et al.* *Nature* 378, 434 (1995).

2. Holliday, R. *Nature* 378, 434 (1995).

3. *The American Public's Attitudes Toward Organ Donation and Transplantation* (Gallup/Partnership for Organ Donation, Boston, 1993).

Why Japan ought to legalize the Pill

Hiromi Maruyama, James H. Raphael and Carl Djerassi

The Japanese Ministry of Health and Welfare is again considering legalization of steroid oral contraceptives. Approval could greatly reduce the incidence of abortion in Japan, which is at least twice that officially reported.

JAPAN is the only industrialized country, except for the Republic of Ireland, where steroid oral contraceptives are illegal¹. Yet Japan was the first industrialized nation after the Second World War where abortion (through the private sector) became the principal method of officially sanctioned birth control; by the mid-1950s, well over a million abortions were recorded annually^{2,3}, and only since the 1960s has the increasing use of other means of birth control — principally condoms and the Ogino calendar rhythm method — resulted in a reduction of abortions, which were officially reported in 1992 to number 410,000 (ref. 3).

High-dose variants of the original pill (G. D. Searle's norethynodrel and Syntex's norethindrone) were approved in Japan in the late 1950s for the treatment of menstrual irregularities and are still used for that purpose. Many Japanese women (estimated at 500,000–800,000)¹ misuse these 'therapeutic' high-dose progestins for contraception. Government approval of 'contraceptive' low-dose pills was denied for nearly three decades on the grounds that insufficient information was available on possible side-effects. When this position became untenable⁴ in the light of thousands of clinical studies throughout the world, Koseisho (the Ministry of Health and Welfare) allowed a safety study in 1986 with 5,000 Japanese volunteers. Contrary to media and industry expectation that the government would legalize the Pill by early 1992, approval was withheld on the grounds that it might lead to reduced condom use at a time when AIDS prevention was considered the highest priority. Many observers¹ attributed this action to other, more complicated reasons that had existed long before Koseisho became belatedly concerned with AIDS in 1991. These include lingering government concern about the potential loosening of sexual mores and Japan's below-replacement birth rate¹, as well as opposition to the Pill by condom-makers and by some physicians licensed to carry out abortions who fear a large loss of income (totalling at least US\$400 million) if an efficacious contraceptive method were widely available. What is more, there has been little pres-

sure from Japanese women for legalization of the Pill — in part because of their ignorance of the greatly reduced risks and non-contraceptive benefits of low-dose oral contraceptives.

Public anxiety about Japan's low birth rate and its rapidly ageing population is understandable. But to use these argu-

standards are even inferior to Japan. Given the overwhelming socioeconomic incentives in modern societies to limit family size, a common aim of all industrialized countries, including Japan, should be to reduce reliance on abortion as a means of birth control. To understand the size of this challenge, it is essential to establish the true rate of abortion in countries where official figures may be suspect. Japan is one such country.

Muramatsu^{7,8} suggested in the 1970s that the number of abortions in Japan was two to four times higher than officially reported. To reach this conclusion, he estimated the numbers of pregnancies and live births among married women that would have occurred without contraceptive use and compared the result with the number of births suppressed by contraceptive practice. Since Muramatsu's study, official abortion figures³ have nearly halved to 410,000 in 1992; Muramatsu's estimate has been neither supported nor rejected in the intervening years. On the basis of the latest statistics and a different approach, we agree with Muramatsu that official Japanese abortion figures are grossly underreported.

Figure 1 summarizes our estimates of abortion rates in Japan for 1992 under several different assumptions about contraceptive use and coital frequency. It is clear that even widely varying parameters do not change the overall conclusion that official figures understate the abortion rate by 100–300 per cent.

To obtain our estimates, we used figures from the 1990 Japanese census and the 1992 *Vital Statistics of Japan* for the numbers of married and unmarried women between the ages of 15 and 44, and the biannual Mainichi surveys^{9,10} for the prevalence of specific contraceptive methods. (For married women, depending on age, the range of condom use is 70–100%, rhythm or basal body temperature method 14–24%, withdrawal 7–13%, interuterine device 1.5–7%, high-dose pill 1.2–6% and other methods 1–2%. Unmarried women report 93% condom use, 18% rhythm method, 5% withdrawal, 1.6% high-dose pill and 3% other methods.) We adopted generally accepted failure rates¹¹ for condoms (12%) and for calendar (20%) and



Japan's way of coming to terms with abortion: a 'graveyard' of commemorative stone statues for aborted babies in a Japanese temple.

ments as a basis for continuing a ban of one of the most widely used methods of contraception is both illogical and unacceptable. Economic and sociocultural factors control the birth rate of a modern population; the quality or use of contraceptive methods is a secondary effect. For instance, the latest data⁵ show that Italy and Spain have the lowest total fertility rates in the world, yet these two countries also have some of the poorest standards of modern birth control methods⁶ in Europe. They have the lowest Pill use, virtually no sterilization and a high dependence (more than 50 per cent) on coitus interruptus or the rhythm method, and in some of these

Picture Credit

withdrawal (18%) methods, and used an admittedly less reliable, recent magazine survey¹² for coital frequency of married couples (depending on the woman's age, typically between once a week and two to three times a month), elsewhere anecdotally reported¹ to be low in Japan, along with Weinstein's¹³ age-related fecundability estimates and, finally, the latest reported³ abortion figure of 410,000. Calculating the number of births expected on the basis of these figures, we attribute the discrepancy between our findings and the annual births reported in *Vital Statistics* to abortion.

Figure 2 summarizes the potential effect of the legalized Pill on the incidence of abortion in Japan. Here we used a relatively conservative set of our 1992 total abortion estimates and compared it with the number of abortions expected if contraceptive use were divided equally between condoms and the low-dose pill. Our calculation indicates that the number of abortions would decrease by roughly 300,000 (28%) in this circumstance.

If we accept that legalization of the Pill by Koseisho¹⁴ would have no effect on the low birth rate in Japan and that sexual mores among the younger Japanese population have been changing without the Pill (even government statistics³ concede that the only age bracket with a significant rise in abortion rate during the past decade is among teenagers), two of the three cited objections to approval of the Pill have little merit. The third — loss of income to abortion providers — clearly should not drive Koseisho policy.

Rather, the time has arrived for the Japanese government and media to acknowledge the persistent high rate of abortion in Japan and to consider its reduction a high priority on medical and social grounds. The addition of oral contraceptives to the meager range of birth control measures available to Japanese

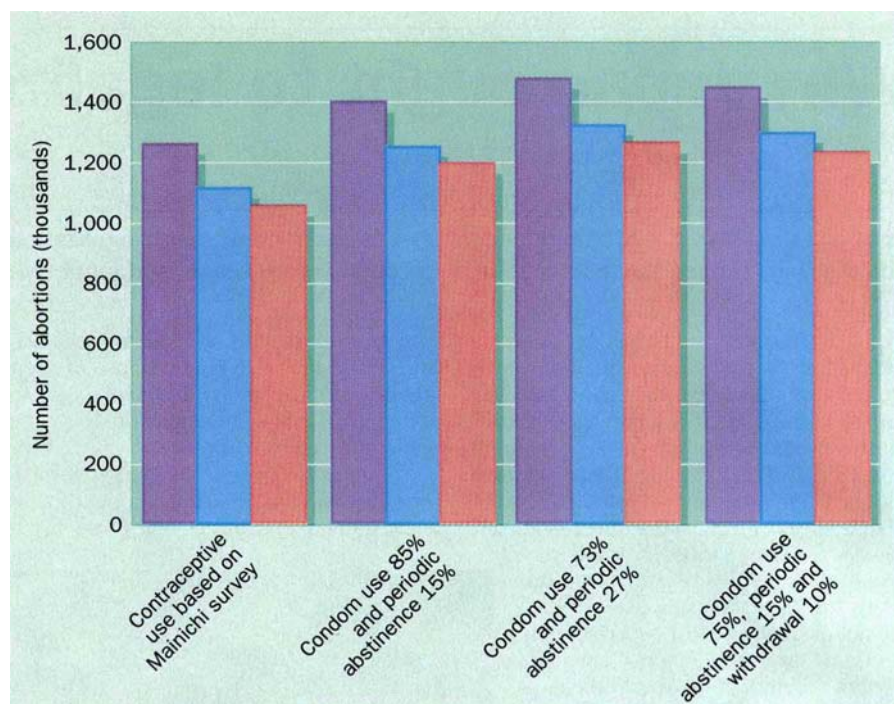


FIG. 1 Number of abortions in 1992 for women aged 15–44 under different assumptions: weighted probability of births based on ref. 13 and the reported coital frequency¹² for both married and unmarried women (purple); replacing weighted probability of births for unmarried women with probability based on coital frequency of twice a month (blue); replacing weighted probability for unmarried women with probability based on coital frequency of once a month for ages 15–19 and twice a month for ages 20–44 (pink). Official number of abortions is 409,069.

citizens would not only enlarge their choices but would also reduce the number of abortions in direct proportion to the number of new users.

But we should not overstate the impact of the Pill alone in reducing abortion in Japan. Given the long history of unfavourable publicity in Japan about the Pill's negative side-effects (owing to the high-dose pill) and the wariness of many women about subjecting themselves to a pharmaceutical programme that artificially regulates their natural hormonal cycle, legalization will almost certainly have to be accompanied by wide-ranging improvements in general sex education, which is notoriously poor in quality. The pharmaceutical industry therefore expects the Japanese market in oral contraceptives to grow only slowly from initially around a million users (including current users of the high-dose pill). Commercial prospects are further limited by the likely division of market share among as many as five manufacturers.

When the Pill is eventually legalized, it should be packaged with a detailed insert of the type required by the US Food and Drug Administration, listing risks

and benefits as well as instructions on use. A sensitively written insert accommodating Japanese cultural practices and up-to-date medical facts would offer an educational bonus far beyond that of ensuring proper contraceptive use. If the leitmotif of government policy became the "The Pill Yes! Abortion No", birth control in Japan would be modernized by 50 years. □

Hiromi Maruyama, James H. Raphael and Carl Djerassi are at the Asia/Pacific Research Center, Stanford University, Stanford, California 94305-6055, USA.

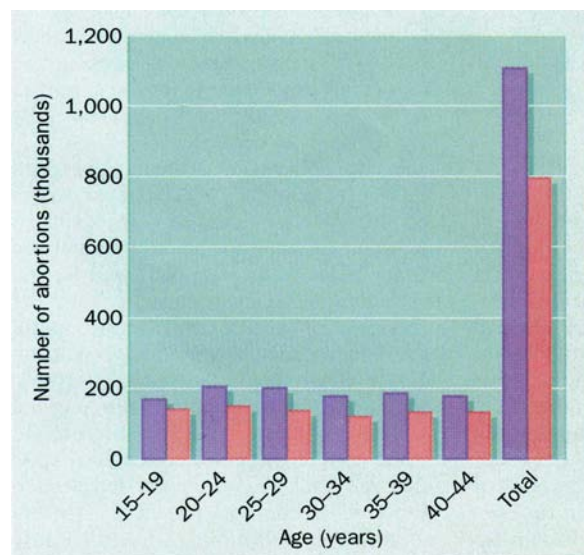


FIG. 2 Estimated effect of Pill use on abortion. Contraceptive use based on Mainichi survey (purple); condom use 50% and Pill use 50% (pink).

1. Jitsukawa, M. & Djerassi, C. *Science* **265**, 1048–1051 (1994).
2. Coleman, S. *Family Planning in Japanese Society: Traditional Birth Control in a Modern Urban Culture* (Princeton University Press, 1983).
3. Koseisho Daijin Kanbo Tokei Johobu [Ministry of Health and Welfare, Minister's Secretariat, Statistics Information Bureau] (ed.) 1992 Yusei Hogo Tokei Hokoku [Eugenic Protection Statistics Report] 20, 44 and 50 (Kosei Tokei Kyokai, 1993).
4. Anderson, A. *Nature* **317**, 760–761 (1985).
5. *World Population Data Sheet* (Population Reference Bureau, Washington DC, 1995).
6. Riphagen, F. E. *Wien. Med. Wochenschr.* **137**(20/21), 488 (1987).
7. Muramatsu, M. *Bull. Inst. publ. Hlth* **22**(4), 228–236 (1973).
8. Muramatsu, M. in *Abortion Research: International Experience* (ed. David, H. P.) 133–136 (Lexington Books, 1974).
9. *Summary of the Twenty-First National Survey on Family Planning* (Population Research Council of the Mainichi Newspapers, Tokyo, 1992).
10. *Summary of the Twenty-Second National Survey on Family Planning* (Population Research Council of the Mainichi Newspapers, Tokyo, 1994).
11. Trussell, J. et al. *Stud. Fam. Plann.* **21**(1), 51–54 (1990).
12. *Nikkei Uerunesu* **12**, 10–36 (1995).
13. Weinstein, M. et al. *Popul. Stud.* **44**, 447–67 (1990).
14. Nathan, R. *Nature Med.* **1**(11), 1115 (1995).

Where has all the old crust gone?

Richard W. Carlson

Geophysicists have come to believe that the Earth was covered with a thick crust more than four billion years ago. New measurements challenge that theory, indicating the importance of crustal destruction in Earth's early history.

VIEWED as a huge chemical factory, the Earth has been trying for 4.5 billion years to separate its chemical constituents into stable, density-stratified layers. Dense iron-nickel alloys, which do not dissolve in the silicate mantle, fell to form the core; volatile compounds were distilled from the interior to make the atmosphere and oceans; and melts of the mantle rose to the surface to form the crust. Most of the core formation and a large fraction of the volatile degassing from the mantle were completed in the first billion years of Earth's history. Evidence for similarly efficient crust formation on the early Earth has been suggested by recent studies of preserved rocks 3.5 to 3.9 billion years old^{1,2}. However, data reported by Vervoort *et al.*³ on page 624 of this issue contradict that evidence, and raise the questions: why did the Earth not form an extensive early crust or, if it did, where has all this old crust gone?

Current ideas of planet formation suggest a very energetic early Earth. Large planetesimals (rock and ice bodies that condensed in the early Solar System) collided violently with the Earth during the latter stages of formation, and transformed their kinetic energy into heat within the growing planet. The sinking of iron alloys to form the core released gravitational potential energy as additional heat. Radioactive elements, perhaps including those that have long-since decayed away, such as aluminium-26 and iron-60 which have half-lives of about a million years, contributed further to heating the early Earth's interior.

Models developed as a result of our visits to the Moon suggest that this energetic formation resulted in planetary interiors that were either largely or totally molten. Upon cooling and crystallization, these 'magma oceans' would have transferred dense material to the core and buoyant crystallizing minerals, such as plagioclase, to the surface. The Moon illustrates this effect clearly in its thick, plagioclase-rich crust; in a mantle showing the complementary chemical signature of plagioclase removal; and in the fact that radiometric dating shows this chemical

separation of crust from mantle on the Moon to have been nearly complete 4.35 to 4.4 billion years ago.

Exploration for ancient terrestrial crust has been intense, but so far the oldest rocks found on the Earth are only 3.96 billion years old⁴, although a few sub-millimetre-sized grains of the particularly sturdy mineral zircon have been found in western Australia that have ages



The 3.54-billion-year-old Ngwane gneiss of Swaziland, southern Africa. This gneiss is composed of feldspar-rich (light coloured bands) and amphibole-rich (dark) material intermixed by metamorphism. The 'stirring' is shown clearly by the feldspar-rich layer folded into an 'S' shape on the right-hand side of the photograph. Metamorphism is probably also responsible for altering neodymium isotope ratios to indicate falsely a thick crust on the early Earth.

(measured by radiometric dating) approaching 4.2 billion years⁵. On the Earth, the lack of preserved ancient crust does not necessarily mean that such crust never existed, because the plate tectonic cycle provides a clear method of crustal destruction through subduction into the deep mantle at convergent plate boundaries, and replacement by new crust produced along ocean ridges.

In seeking information on Earth's first crust, one can remove the effects of more than 3 billion years of plate tectonics by

looking at the ratios of different radiogenic isotopes in the oldest preserved crustal rocks at the time of their formation. The chemical changes associated with crust formation result in abundance ratios of parent to daughter elements that differ between the crust, its complementary melt-depleted mantle and mantle unaffected by melt removal. Over time, these distinct parent/daughter ratios produce different daughter-element isotopic compositions. This is demonstrated well in the volcanic rocks erupted today along the Earth's ocean ridges, which show very clearly that the mantle providing these magmas is a complementary residue to the continental crust.

Three of the radioactive decay systems that have been used for this purpose are rubidium-strontium, samarium-neodymium and lutetium-hafnium. Neodymium data taken in the early 1980s (ref. 6) from crustal rocks 2.7 to 3.8 billion years old showed clearly that, 3.8 billion years ago, the mantle had already been depleted by extraction of crust. The deviation of these data from those expected for an undepleted mantle was not large, however, and failed to provide evidence for the existence of large volumes of crust produced by magma ocean crystallization.

More recent data, particularly for the oldest preserved crustal rocks^{1,2}, show much wider initial neodymium isotope variation, implying the existence of large volumes of crust in the Hadean era (the geological period between Earth's formation and creation of the oldest preserved crust).

Calculating initial isotopic compositions of old rocks, however, relies on the unverifiable assumption that the parent/daughter ratio of the sample has remained constant since rock formation. In general, the assumption is a good one, but crust that has floated around Earth's surface for billions of years has had many opportunities for collisions with other crustal blocks, which can lead to metamorphism of the constituent rocks (see figure). If that changes the parent/daughter ratio of a sample of old crust long after

rock formation, the initial isotopic composition will be obscured. Because of the ease with which the rubidium/strontium ratio of a rock can be changed by metamorphism, attempts to study crustal development by examination of that radiometric system have all but been abandoned. Samarium and neodymium, being neighbouring lanthanide elements and insoluble in most metamorphic fluids, are more immune than rubidium to migration during metamorphism, but the question that must be faced is, are they absolutely immobile?

To avoid this problem, Vervoort *et al.* examined the lutetium–hafnium radiometric system in the mineral zircon. Zircon is well known to geochronologists because it incorporates high quantities of uranium but almost no lead during its formation, and so is extremely useful for precise uranium–lead geochronology. Furthermore, it is a very stable mineral during metamorphic processes. For lutetium–hafnium studies, zircon offers the additional advantage that hafnium is an abundant element whereas its parent lutetium is present only in trace amounts. Consequently, the lutetium/hafnium ratio of zircon is near zero, necessitating only minor corrections to measured hafnium isotopic compositions.

Vervoort and colleagues show that the initial hafnium isotope compositions of zircons from Greenland rocks 3.6 to 3.8 billion years old indicate a mantle source that was mildly depleted by melt removal in the Hadean era. However, the range of initial neodymium isotope composition of the same rocks is much larger than is compatible with the observed variation in hafnium, implying that the more extreme values of neodymium composition observed in ancient continental rocks are the result of metamorphic perturbations, and not an indication of large amounts of ancient crust.

We are faced with a troubling lack of evidence for large volumes of Hadean crust on Earth. Perhaps we are looking for the wrong type of crust? The lunar example of a plagioclase-rich flotation crust is probably not appropriate for the Earth, where much greater interior pressures promote crystallization of the dense mineral garnet, which would consume the aluminium necessary for plagioclase formation. Another explanation, but by no means the only one, is that vigorous convection in the early mantle was particularly good at mixing away evidence of earlier

crust–mantle differentiation. If this is the case, the Earth's surface in the Hadean may have been covered with a thin and transient basaltic crust, much like the current situation in the ocean basins where new crust formed along ocean ridges survives at the surface for, at most, a couple of hundred million years before it plunges back into the mantle in a subduction zone. The continent–ocean distinction, unique

to the Earth among the terrestrial planets, may not be a primary feature of Earth, but rather one reflecting the losing battle of plate tectonics to stir back the 'slag' separating from the mantle 'foundry'.

Richard W. Carlson is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.

CLIMATE CHANGE

Doctors watch the forecasts

William A. Sprigg

BLAME the weather. From the Black Death in 14th century Europe to the possible spread of tropical diseases into temperate regions in the next few decades, the climate is a powerful controlling factor for many aspects of human disease. That was the message of an 'All Union' session during the American Geophysical Union's meeting in December*, which explored the connection between climate and health. The timing of outbreaks and the spread of disease are affected by temperature, rainfall, sunshine and ocean currents, which can limit the range of carriers or alter the whole complicated ecology of a region. Understanding these links may prove vital in fighting disease.

Malaria, cholera and dengue fever are among the feared diseases that are linked to changes in weather through 'vectors' (pathogen carriers such as ticks, mice and mosquitoes) and vertebrate reservoirs. The climate–disease link is of increasing concern because of recent outbreaks of emerging infectious diseases, such as Ebola in Africa, and those increasing in incidence or in geographical range, such as cholera in South America and Lyme disease in the United States. Further apprehension is aroused by the possibility of global warming, which could alter traditional ranges of disease vectors.

On the other hand, new tools for combating disease are becoming available, including satellites and weather-radar, global data networks, climate analysis and prediction, and increasing international cooperation in environmental research. For example, useful advances have been made in understanding weather changes associated with the El Niño/Southern Oscillation phenomenon (ENSO), in which warm ocean waters move from the western Pacific to the central and eastern Pacific, affecting global atmospheric circulation and causing dramatic changes in temperature and precipitation. For some countries affected by ENSO, new predictive capabilities are already being exploit-

ed for agricultural purposes. In recognition of such progress, an international agreement was reached in November 1995 to establish a research institute for climate prediction and its applications.

A recent El Niño event appears to be closely related to the 1991–1992 cholera outbreak in Central and South America (Rita Colwell, Univ. Maryland). Plankton blooms that occur in warm water can sequester the bacterium *Vibrio cholerae*, which is then ingested by humans while bathing or using untreated water for drinking or preparing food. Several malaria outbreaks in the past three decades also appear to be associated with ENSO events and the consequent rainfall increases.

The 1992–1993 El Niño may have provided conditions that led to the deadly outbreak of Hantavirus pulmonary syndrome in the southwestern United States in 1993 (James Gosz, Univ. New Mexico). Among many complex factors, a six-year drought stressed vegetation and reduced the populations of predators of the common deer-mouse, a vector for the transmission of Hantavirus. The reappearance of rain produced abundant food, and before their natural predators returned mouse populations soared, making contact with humans inevitable. A similar sequence of events may have contributed to the Black Death and other 14th century plague outbreaks in Europe and China (Kevin Pang, La Canada, California).

Climate variability can help spread infectious diseases in other ways. Human migration driven by drought-induced crop failures can carry a previously isolated infection to other populations. Travellers can carry pathogens into a new, climatically receptive environment. Warmer temperatures can influence specific vectors by increasing their poleward distribution, their source of food and their own lifespan, accelerating larval development and decreasing the time from infection of the vector until it is able to transmit the disease (Robert Shope, Univ. Texas Medical Branch). Vector distribution with altitude up mountain slopes will also be affected (Paul Epstein, Harvard Medical School),

1. Jacobsen, S. B. & Dyrnek, R. F. *J. geophys. Res.* **93**, 338–354 (1988).
2. Bowring, S. A. & Housh, T. B. *Science* **269**, 1535–1540 (1995).
3. Vervoort, J. D., Patchett, P. J., Gehrels, G. E. & Nutman, A. P. *Nature* **379**, 624–627 (1996).
4. Bowring, S. A., King, J. E., Housh, T. B., Iaschsen, C. E. & Podosek, F. A. *Nature* **340**, 222–225 (1989).
5. Froude, D. O. *et al.* *Nature* **304**, 616–618 (1983).
6. Shirey, S. B. & Hanson, G. N. *Geochim. cosmochim. Acta* **50**, 2631–2651 (1986).

*American Geophysical Union Fall Meeting, San Francisco, USA, 11–15 December 1995.

and may be useful as an early warning of wider geographical spread.

Environmental conditions are always changing. The frequency and intensity of El Niño events have increased since the late 1970s and precipitation has decreased in the subtropics and increased over land in the Northern Hemisphere at mid- and high latitudes, which is also where the greatest surface warming has occurred, realized in night time temperature (Thomas Karl, National Oceanic and Atmospheric Administration). Monitoring for such changes can help target disease vectors for control, prepare local health officials for stepped-up surveillance and increase public awareness. Ecosystem and climate monitoring have been recommended to aid surveillance and prediction of epidemics of childhood bacterial meningitis in Africa (Stephen Morse, Rockefeller Univ.).

A set of tools that could be exploited are the satellite sensors that have been used to map frost-prone areas, define forest boundaries and locate plankton blooms. Such abilities may be useful in identifying the distribution of coastal waters that harbour *V. cholerae*, for example, or the preferred ranges of the mosquitoes responsible for encephalitis or of the ticks that transmit Lyme disease. With little effort, probability distributions of favourable environmental conditions for disease vectors that are based on accurate descriptions of current climate can be routinely available. Satellite projects launched for other purposes — measuring soil wetness, ocean colour (for plankton), or tropical rainfall — can help to explore the connections between climate and health.

Scientific study of climate–health connections has been slow to develop. Advances to date have generally come from lone investigators lacking the comprehensive, multidisciplinary interaction that has so benefited other climate research, such as the investigation of El Niño through the Tropical Ocean and Global Atmosphere programme. Such broad-based studies are needed because disease vectors are affected by so many components of the ecosystem.

Considering the wealth of technical resources available, the will and the mission of many organizations to assist, and the scientific imperative, it is time to develop an appropriate international research strategy. The links between climate and the outbreak and spread of infectious disease are well documented in some cases, suspected in more, and undoubtedly will be discovered in others. It is indeed important to know when, and to what extent, these links exist.

William A. Sprigg is in the National Research Council, National Academy of Sciences, Washington DC 20418, USA.

Soldier production under threat

Deborah M. Gordon

SOMETIMES a theory is so appealing that it is widely accepted even though there are few data to confirm it. This is what happened with Oster and Wilson's theory of adaptive caste distributions in social insects¹, in which they predicted that the distribution of specialized workers in a colony should be optimized. On page 630 of this issue², however, Passera *et al.* demonstrate a remarkable shift in caste ratios in response to environmental conditions.

In 1978, Oster and Wilson published one of the first applications of dynamic programming in evolutionary biology. Their argument, based on the example of polymorphic ant species in which adult workers differ in body size, was that colonies of social insects might be composed of specialized individuals, or 'castes', each highly efficient in the performance of a particular task. Therefore, the argument continues, natural selection should produce the optimal distribution of workers into various castes. Like a factory assembly line, the numbers of workers in each caste would maximize the colony output, each caste contributing to the production of new colony members, in a safe nest with the best possible food supply. If body size is correlated with task efficiency, then the optimal distribution of body sizes will be the one that makes the colony most effective at performing all its tasks.

According to Oster and Wilson's theory, caste distributions should be fixed once a colony reaches maturity, and should be adaptive on the evolutionary timescale, over many generations. This prediction has rarely been tested; one study suggests that factors other than task efficiency drive selection on the distribution of body sizes³. Most empirical studies of caste distributions have addressed adaptation on the much shorter ecological timescale, within the lifetime of a colony. It is not clear that these studies provide a test of the second step in Oster and Wilson's argument. Most of them found no evidence that colonies adjust the production of workers of a given body size to ecological conditions⁴.

But now Passera *et al.*² provide evidence for an adjustment of caste distributions in response to the colony's environment. In the ant genus *Pheidole*,

adult workers are either small 'minors' or large 'majors'. In some species, including *P. pallidula*, the one studied here, the majors act to defend the colony, earning them the title of 'soldiers'. Passera *et al.* exposed ants of one colony to those of another, using a wire mesh that allowed legs or antennae through but prevented direct attacks. Such contact led colonies to produce more soldiers, the implication being that the colony is thereby better equipped to defend itself from conspecific neighbours.

The production of more large workers to defend a colony is a new example of an



Ant attack — aided by minor caste members from the same colony, a soldier *Pheidole pallidula* gets to grips with a soldier worker from another colony.

inducible defence. Another case is found in certain bryozoans, which produce spines in the presence of predators. Such structural inducible defences are rare in animals⁵, however, and models of the evolution of phenotypic plasticity show why: defences like these are optimal only in a variable environment that supplies very dependable cues to the presence of danger. Moreover, the cost of producing new body parts (or, in the ant-colony case, new bodies) and the cost of the delay before defences are available, can outweigh the benefits that inducible defences provide⁶. This kind of reasoning led Oster and Wilson⁷ to suggest that the alteration of soldier/worker ratios should be rare in ants, even on the evolutionary timescale.

A colony could bolster its defences without producing more soldiers. Ants previously engaged in other tasks could switch to help with defence, and short-term adjustments of this sort are well known in other species of *Pheidole*: when ants of one caste are removed, ants of the other caste change tasks to compensate for their missing nestmates⁸. Making more soldiers from scratch is a much

L. Passera

slower way for a colony to protect itself; many weeks may elapse between the larval stage, when body size is determined, and the emergence of an adult. The production of new, large workers is costly as well as slow. Such costs include the energy differential needed to produce a large worker, and the losses the colony might suffer from enemy attack while it waits for more soldiers to develop. Benefits depend on how much better majors are than minors at defending the colony, and whether the presence of ants from another colony provides a reliable signal, with sufficient warning time, of future aggression.

Pheidole is a large and widespread genus, showing a remarkable diversity of behaviour. In many species, members of the major caste mill seeds or act as storage vessels for food, and do not carry out defensive duties⁸. Thus *P. pallidula* seems to be an unusual member of this evolutionarily labile genus. Rather than confirming that adaptive caste distributions are the rule, the new observations raise

intriguing questions about this exceptional species. Is its ecology unusual among the *Pheidole*? Does a *P. pallidula* colony take many weeks to carry out an attack on another? Are the majors of this species more ferocious than those of other *Pheidole* species? Are the minors more timid? Passera and colleagues' result offers a new opportunity to investigate the evolutionary ecology of inducible defences. □

Deborah M. Gordon is in the Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA.

1. Oster, G. & Wilson, E. O. *Caste and Ecology in Social Insects* (Princeton Univ. Press, 1978).
2. Passera, L., Roncin, E., Kaufmann, B. & Keller, L. *Nature* **379**, 630–631 (1996).
3. Beshers, S. N. & Traniello, J. F. A. *Ecology* **75**, 763–775 (1994).
4. Schmid-Hempel, P. *J. evol. Biol.* **5**, 1–12 (1992).
5. Adler, F. R. & Harvell, C. D. *Trends Ecol. Evol.* **5**, 407–410 (1990).
6. Levins, R. *Evolution in Changing Environments* (Princeton Univ. Press, 1968).
7. Oster, G. & Wilson, E. O. *Caste and Ecology in Social Insects* Ch. 6 (Princeton Univ. Press, 1978).
8. Wilson, E. O. *Behav. Ecol. Sociobiol.* **16**, 89–98 (1984).

NEUROBIOLOGY

The information superflyway

Rodney J. Douglas and Kevan A. C. Martin

It is a truth universally acknowledged that, to survive, living organisms are in need of accurate information about their world. This information is collected by a dazzling array of specialized receptors that transduce the stimulus energy, whether it be light, pressure, tension, heat, chemical or gravity, or whatever, into the internal signals of the nervous system.

Two papers published elsewhere in this issue^{1,2} examine related aspects of this early processing in the visual system of blowflies. The neurons under scrutiny were the photoreceptors and the large monopolar cells (LMCs) in the compound eye, and two of the three classes of motion-sensitive neurons of the lobular plate, a further stage in the visual pathway. All are non-spiking neurons, so-called because they do not communicate with other neurons by nerve impulses, but instead transmit their messages across synapses as analogue signals coded in the form of graded voltages.

De Ruyter van Steveninck and Laughlin (page 642)¹ recorded the electrical activity of the photoreceptors in response to fluctuations in light intensity, and used Shannon's information theory to estimate that the rate of information capture was 1,000 bits s⁻¹. The LMCs are second-order neurons that receive convergent and thus redundant information from six photoreceptors. They achieve a capacity of 1,650 bits s⁻¹. These numbers are striking

because they are much larger than the maximum rate of about 300 bits s⁻¹ previously estimated for the third class of neurons in the lobular plate, which do communicate by nerve impulses³.

One of the important ways in which the fly has to use this rich flood of information is to determine its motion relative to the world. In studying the neurons that are involved in this computation of motion, Haag and Borst (page 639)² discovered that transmission of visual signals along the dendrites of one subclass of motion-sensitive neurons in the lobular plate of the blowfly is enhanced by a fast sodium current that acts selectively to amplify incoming signals of high frequencies. Their experiments provide direct evidence for the importance of active mechanisms for the analogue processing of information that is carried out within all neurons.

These two papers, coming from different conceptual frameworks, illuminate a central question about the nature of the encoding of the world by a nervous system. What is the organism's strategy for balancing the need for extensive knowledge with the need for specific and immediate action? The tools of information theory used by de Ruyter van Steveninck and Laughlin were developed by Shannon and Weaver⁴ for a very specific purpose, which was to calculate information transmission across a channel in a relatively well defined and simple physical system,

such as a telephone network. In the case of the nervous system, the elements and processes, and their interactions, are often poorly understood. Consequently, information theoretic analyses of neural processing are confronted with substantial experimental difficulties, and must make bold, even extreme, simplifying assumptions. Among the most successful projects have been those that have addressed quite simple questions about the amount of information coded by spiking neurons.

From the time that Adrian⁵ made the fundamental observation that the intensity of a stimulus is coded in the frequency of identical nerve impulses, it has been appreciated that the only data about the world to which the central nervous system has access are the absolute time of occurrence, or the time of occurrence of one impulse relative to other impulses. The critical question that information theorists have posed is this: what is the maximum amount of information that such a train of impulses could convey about the stimulus that it reflects? Notice that the question is posed in terms of the 'ideal observer', a platonic being who has the means to extract all the available information in the impulse train.

Bialek *et al.*³ used optimization methods to construct the best 'ideal observer' for spiking neurons in the lobular plate of the blowfly and estimated that they could transmit information at a rate of about 300 bits s⁻¹. De Ruyter van Steveninck and Laughlin¹ have now used a related mathematical method to show that the rate of information transmission from photoreceptors to the next non-spiking neurons in the circuit is five times higher. Assuming that the 10⁷ photoreceptors of the primate retina have equivalent performance, then the retinal receptors would seem to encode about a gigabyte per second — a data rate that would fill the hard disk of the latest personal computer in a second. Even for the few thousand photoreceptors of the compound eye of the fly, the potential information load is close to 0.5 megabytes per second.

This performance by non-spiking neurons is, of course, striking. But the reasons why the frequency-coding method used by spiking neurons is relatively slow and imprecise were pointed out by von Neumann almost 40 years ago⁶: non-spiking neurons can convey information in analogue form up to the frequency limits of their voltage fluctuations, whereas spiking neurons have to code the information by varying the interval between successive impulses, the interval being typically tens of milliseconds. Apparently, spiking neurons use about the least optimal means of coding.

However, the brain's higher visual centres are not television screens that receive a high-fidelity transmission of the visual

image, which then has to be viewed and interpreted by some observer. Instead, as Barlow has emphasized⁷, information from the photoreceptors appears to be processed in subsequent stages so as to use the fewest possible number of active neurons to achieve as complete a representation of the stimulus as is required for an appropriate behaviour. Such selective representations have the advantage of reducing considerably the volume of information that has to be propagated by a single neuron in a given time. Haag and Borst's experimental *tour de force* shows that some neurons use special mechanisms to transmit some of the information selected for the representation of motion.

The mechanism of active dendritic amplification by sodium currents, discovered by Haag and Borst, provides a means for boosting transmission of the high-frequency components of the incoming motion signal. But the amplification is not simply linear. Because activation of the sodium currents is voltage dependent, the degree of amplification depends on the average dendritic depolarization. So the amplifier could possibly be switched on or off by the character of the input signal, or some control neuron. Moreover, the dynamics of the active process places limits on the fidelity of the signal transmission. Their results show that although the neuron can transmit the frequency of a sinusoidally modulated visual stimulus, the nonlinear amplification introduces significant distortions in the high-fidelity signal that is transmitted from the photoreceptors. These motion-sensitive neurons are not ideal observers.

Clearly, experimental work based on information theoretic analyses is a fertile area that could at least characterize the information loads faced by the nervous system and the transmission rates of this information along the processing paths. Such analyses may also lead to insights about how information is transformed into structural and functional properties of the neuronal circuits that represent an organism's knowledge of the world. As these two papers show, however, this is not an area for the faint-hearted. □

Rodney J. Douglas and Kevan A. C. Martin are in the Institute of Neuroinformatics, ETH and University of Zürich, Gloriastrasse 32, Zürich 8006, Switzerland.

Iron grip on export production

Alan Longhurst

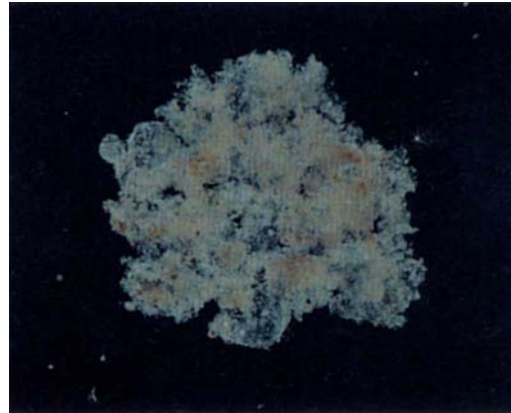
SINGLE-CELLED microscopic plants living near the ocean surface take up carbon, and some of this carbon sinks out of surface waters and into the abyss below. This process, called 'export production', is an important link in the global carbon cycle and perhaps in climate change, but in many cases we do not understand what controls its rate. Two *Nature* papers give us part of the answer: over some of the

observations of Kumar *et al.*² convincingly support the late John Martin's suggestion³ that during the Last Glacial Maximum there were higher fluxes of iron-rich aeolian (wind-borne) dust than there are now, and that they induced increased plant growth and sedimentation of organic carbon in the Southern Ocean, at present an HNLC area having very low fluxes of aeolian dust. Martin suggested that this process contributed to the reduction of atmospheric carbon dioxide concentrations during the period. Kumar *et al.* examined marine sediments for three proxies (the ratios of protactinium-231 and beryllium-10 to thorium-230, as well as the burial rate of uranium precipitates produced *in situ*) of productivity and export production, and demonstrated that these ideas are substantially correct, though with some change of emphasis.

Increased export production was restricted to the subantarctic zone, and was associated with evidence for strongly enhanced transport of aeolian dust from Patagonia. Uranium burial rates indicate organic carbon export comparable to that observed in present-day coastal upwelling centres. They show that increased iron concentrations lead to increased export, and that the ceiling on this

export — known from other modelling studies that assume there is enough iron to encourage production and all the nitrate is used up — is about 15 to 30 p.p.m. (parts per million) CO₂ drawdown, to be compared with around 80 p.p.m. drawdown observed for the glacial atmosphere.

Coale and colleagues address limitation by iron in a very different environment, today's equatorial Pacific Ocean near 140° W, where aeolian deposition at the sea surface is also low⁴. Earlier reports from their field studies⁵ described how upwelling at the Equator maintains locally high levels of nitrate. Coale *et al.* now propose that iron from shallow hydrothermal sites in the western Pacific is entrained in the ribbon-like undercurrent which lies below the Equator, 200 m deep, 200 km wide and approaching 10,000 km long, and that upward mixing of iron and nitrate from the high-velocity core at 200 m supports a plankton bloom near the surface. However, the rates of supply of iron and nitrate are so unbalanced that only 20 per cent of the nitrate can be used by plants. This is consistent with the earlier report⁵ that the iron-nitrate imbalance at the Equator causes suboptimal photosynthetic performance characteristic of



R. S. Lampitt

'Marine snow' particle (an aggregation of dead phytoplankton, zooplankton and faecal matter) about 5 mm across, collected from the seabed at a depth of 1,400 m on the northeastern Atlantic continental slope. The brown regions of the particle are rich in plant pigment, indicating that it sank quickly to the seabed before much degradation took place — one way that carbon can be carried away from the surface.

ocean, the controlling factor is iron.

During the early evolution of plants the geochemical balance of our planet was thrown out of equilibrium as carbon dioxide was progressively removed from the primitive atmosphere by photosynthesis. Most of the carbon is sequestered in biogenic marine carbonates, and to a smaller extent in coal and shales. From this observation stems an urgent need to understand how marine plants, dominated by the single-celled phytoplankton of the open ocean, affect atmospheric CO₂ during natural climate changes. We may then hope to predict their response to even higher CO₂ levels than those already generated by forest clearance, cement production and fossil-fuel burning. The results of Coale *et al.*¹ (reported on page 621 of this issue) and Kumar *et al.*² bring us closer to an understanding of the limitation on phytoplankton growth, and on sinking of carbon to the sediments, in those large regions of the ocean where the classical limiting macronutrient nitrate (NO₃) remains in excess.

Both papers support the growing body of evidence for a critical lack of available iron in some of these HNLC (high nutrient, low chlorophyll) ocean areas. The

1. de Ruyter van Steveninck, R. R. & Laughlin, S. B. *Nature* **379**, 642–645 (1996).
2. Haag, J. & Borst, A. *Nature* **379**, 639–641 (1996).
3. Blalock, W., Rieke, F., de Ruyter van Steveninck, R. R. & Warland, D. *Science* **252**, 1854–1857 (1991).
4. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (Univ. Illinois Press, Urbana, 1949).
5. Adrian, E. D. *The Basis of Sensation* (Christophers, London, 1928).
6. von Neumann, J. *The Computer and the Brain* (Yale Univ. Press, New Haven, 1958).
7. Barlow, H. B. *Perception* **1**, 371–394 (1972).

iron-limited cells in this HNLC area.

Reports are also now emerging^{6,7} from the recent field simulation (IronEx-II) at the 10-km scale of the effect of aeolian iron deposition in the eastern tropical Pacific HNLC region. During IronEx-II, repeated enrichment (to 1–2 nanomolar) of surface water with iron sulphate solution pumped into the wake of a research ship led to an order of magnitude increase in plant mass, a 5 micromolar decrease in nitrate concentration, and a 100 p.p.m. decrease in the partial pressure of CO₂.

IronEx-II also confirmed what had been suspected from many somewhat unsatisfactory *in vitro* incubation experiments, that it is the larger phytoplankton cells, principally diatoms, that respond when iron is added to surface water from HNLC areas. These clear surface waters are almost exclusively populated by pico- and nanoplankton, which have little potential to export organic material to the sediments — largely because they are too small to sink very far before being degraded — and whose productivity is based on recycled ammonia rather than on available nitrate. Modification of such an ecosystem by the growth of a second, diatom–copepod community (the copepods are zooplankton that can graze on the diatoms) using available nitrate will not only increase the overall production rate and biomass, but will also enhance export production because the larger cells (which often aggregate) and faecal pellets will sink rapidly into the interior of the ocean. Though there were no direct estimates of this export using sediment trap data, such an enhancement is thought to have occurred during IronEx-II.

An important conclusion should be drawn from all this as we continue our attempts to predict the response of phytoplankton to global-scale changes in oceanic and atmospheric circulation, and the sequestration of atmospheric CO₂ into the interior of the ocean and deep-sea sediments. It will no longer be sufficient to model algal growth as a response to physics and the supply of NO₃ to the near-surface 'photic zone': we must also include iron transport and supply. Only thus can we expect to quantify the response of the ocean to changing global climate regimes, or to predict where iron-limitation may occur in the larger regions

currently characterized by nitrate-limitation. Global maps of aeolian dust deposition⁴, phytoplankton growth rates⁸ and biomass⁹ show that this critical task will be complex, for there is still much to be learned about the present situation. Paradoxically, this is perhaps especially true in the much-studied northeastern Pacific,

where herbivore constraints on plant growth in the presence of non-limiting iron and nitrate may be a special case¹⁰. □

Alan Longhurst was formerly in the Biological Oceanography Division, Bedford Institute of Oceanography, Dartmouth B2Y 4A2, Nova Scotia, Canada.

ASTROPHYSICS

Harassed galaxies blush

Bob Joseph

ONE of the major insights in astrophysics gained during the past decade is the realization that interactions between galaxies play a considerable role in the evolution of many and probably most field galaxies, that is, those outside clusters. Following the classic numerical simulation of galaxy interactions by the Toomre brothers¹ in 1972, simulations have become highly detailed and sophisticated as a result of the enormous advances in computational power and techniques. However, there is one sort of interaction that has so far received less attention in numerical simulations, and that is when the relative speed of the two galaxies is large compared with their rotational and internal speeds, as is the case for galaxies in clusters. On page 613 of this issue², Moore *et al.* briefly describe the results of a new set of simulations of multiple high-speed interactions, which they argue induce profound evolutionary transformations for galaxies in clusters.

They are hoping to understand why nearby clusters are dominated by very red elliptical and lenticular galaxies, whereas clusters observed earlier in their evolution, with redshifts >0.2 , have large populations of blue galaxies. This strong evolution, called the Butcher–Oemler effect^{3,4}, is not observed in field galaxies, and so must be an effect of the cluster environment.

Several years ago Lavery and Henry⁵ reported evidence that the blue galaxies were undergoing galaxy–galaxy interactions, and they suggested that interaction-induced bursts of star formation had produced the young hot stars which gave these galaxies their blue colours. Once the starbursts had consumed all the available gas there could be little further star formation and the galaxies would exhibit the redder colours of cooler stars. Moore *et al.* are following up this idea by modelling the high-speed interactions experienced by galaxies in clusters.

High-speed encounters seldom, if ever, occur between field galaxies. Moore *et al.* realized that, in contrast, clusters should exhibit the effects of many such high-speed encounters, since the galaxy density is much greater than it is in the field, and the galaxies are on bound orbits in the

gravitational potential of the cluster. They describe this process of repeated high-speed interactions as “galaxy harassment”. (I suppose it had to happen sooner or later. In their pioneering numerical simulations of interactions between galaxies, Toomre and Toomre referred to the two galaxies in the interaction as “victim” and “aggressor”. Any day now we may expect to see mergers between galaxies described as “galaxy abuse”.)

Moore and colleagues display the “disturbed” morphologies of the harassed galaxy, which appear in their simulations after one or two encounters, and argue that the correspondence with morphologies seen in recent high resolution images of Butcher–Oemler clusters (dominated by blue galaxies) is striking. Similarly, they compare the properties of a simulated galaxy after several high-speed encounters with those of the red galaxy population in nearby clusters, and they argue that the Butcher–Oemler effect can be understood as the consequence of the morphological transformations induced by this continued harassment.

As intriguing as the results of Moore *et al.* are, a number of questions arise which are not addressed in their brief paper. Numerical simulations are always an approximation to reality, and one of the approximations Moore *et al.* make is to treat all the “harassing” galaxies in the cluster as ‘softened’ point masses, that is, rigid potential wells. When these galaxies, rather than just the harassed galaxy, are properly modelled as *N*-body systems, the resulting galaxy morphologies could be quite different. They also ignore the effects of star formation on the evolution of the gas in their simulations. Mass in gas is dissipative and loses energy, whereas mass in stars does not, and it is not unlikely that the extreme radial shrinkage of gas they find and the resulting dwarf galaxies are artefacts of this approximation.

Does galaxy harassment as described in these numerical simulations solve the mystery of the Butcher–Oemler effect? In my view it does not. For most astronomers, numerical simulations are something of a black box. Until similar results are obtained by others using different numeri-

- Coale, K. H., Fitzwater, S. E., Gordon, R. M., Johnson, K. S. & Barber, R. T. *Nature* **379**, 621–624 (1996).
- Kumar, N. *et al.* *Nature* **378**, 675–680 (1995).
- Martin, J. *Paleoceanography* **5**, 1–13 (1990).
- Duce, R. A. & Tindale, N. W. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
- Lindley, S. T., Bidigare, R. R. & Barber, R. T. *Deep-Sea Res.* **42**, 441–462 (1995).
- Coale, K. *et al.* *EOS (Ocean Sciences Meeting)* **154** (1996).
- Watson, A. J., Cooper, D. J. & Van Scoy, K. A. *EOS (Ocean Sciences Meeting)* **177** (1996).
- Longhurst, A. R., Sathyendranath, S., Platt, T. & Caverhill, C. J. *Plankton Res.* **17**, 1245–1271 (1995).
- Feldman, G. *et al.* *EOS* **70**, 634–641 (1989).
- Banase, K. *Limnol. Oceanogr.* **35**, 772–775 (1990).

cal codes, and some of the simplifications mentioned above are adequately addressed, I suspect that many will find it difficult to lend too much credence to the results of any one simulation. The significance of this work for me is that it provides the first, admittedly crude, numerical simulations qualitatively supporting Lavery and Henry's idea that interactions are the key to understanding the recent evolution of cluster galaxies. This is a step forward, since it had previously been easy to dismiss categorically the possible importance of interactions with the superficial observation that high-speed interactions in clusters don't 'do' anything.

The shift towards an appreciation of the importance of interactions in the evolution of field galaxies resulted from a wonderful interplay between numerical simulations and observational studies. I hope these new results stimulate further numerical simula-

tions of increasing sophistication. Concurrently, there will certainly be further observational studies of galaxies in clusters using the newly-available tools provided by both ground-based and space-borne telescopes. With continuing interplay between simulations and observations, I expect an understanding of the Butcher-Oemler effect to emerge as one of the astrophysical achievements of the current decade. □

Bob Joseph is at the Institute for Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA.

1. Toomre, A. & Toomre, J. *Astrophys. J.* **178**, 623–666 (1972).
2. Moore, B., Katz, N., Lake, G., Dressler, A. & Oemler, A. Jr *Nature* **379**, 613–616 (1996).
3. Butcher, H. & Oemler, A. *Astrophys. J.* **219**, 18–33 (1978).
4. Butcher, H. & Oemler, A. *Astrophys. J.* **285**, 426–438 (1984).
5. Lavery, R. J. & Henry, J. P. *Astrophys. J.* **330**, 596–608 (1988).

MOLECULAR EVOLUTION

Digging up the roots of life

Arne Ø. Mooers and Rosemary J. Redfield

WHEN did life arise? What did it look like? These questions were once the province of theologians, but in recent years palaeontologists have been better guides. Molecular systematists have now joined the game, and the latest molecular evidence, published last month in *Science*¹, points to a date almost 2 billion years (Gyr) younger than that implied by the fossil record.

Until now, the fossil evidence that life arose and bacteria diversified more than 3.5 billion years ago, soon after the Earth first developed a crust and ocean, has been virtually uncontested. This evidence consists of both structurally complex microfossils with a striking resemblance to modern photosynthetic cyanobacteria, and stromatolites, layered structures resembling modern cyanobacterial mats².

Searching for deep roots in the tree of life has lately become a hot topic for molecular systematists. Comparisons of slowly evolving RNA and protein sequences have identified bacteria, Archaea ('archaeobacteria') and eukaryotes as the three main domains of life, and it seems more and more likely that the first divergence was between bacteria and (Archaea + eukaryotes), with all of the modern bacterial groups diversifying after this initial split³. Although most deep phylogenetic trees have been uncalibrated (that is, based on sequence divergence, not actual time), the resemblance of the oldest fossils to modern bacteria implies that this initial divergence occurred more than 3.5 Gyr ago. But Doolittle and colleagues¹ now present an analysis placing the first branch much more recently, at about 1.8

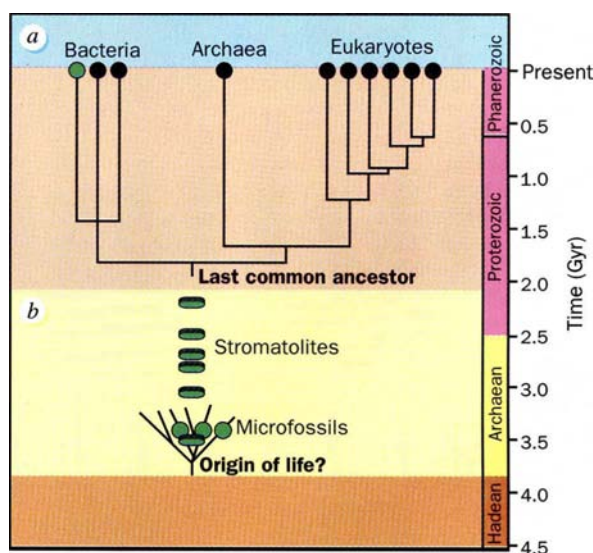
Gyr ago. This means that our early history might be in need of serious re-evaluation.

Although squeezing evidence from billion-year-old stones is difficult business, the evidence for 3.5-Gyr-old bacteria seems strong. Because many reported 'fossils' turn out to be bubbles of non-biogenic matter, contaminants from later eras or even spores floating about the laboratory⁴, micropalaeontologists have agreed upon a set of stringent criteria addressing provenance and protocol, as well as biogenic origin. These include morphological relationships to modern groups, and evidence of plausible morphological variation and developmental stages. Based on these criteria, there are now two sites that have produced convincing fossils reliably dated at about 3.5 Gyr old. These sites have produced a wide array of forms, often attributed to different families and genera of modern cyanobacteria, a group with complex morphologies allowing tentative assignment to known taxa. Micropalaeontologists go so far as to speak of diversified communities and ecosystems represented in the Archaean rocks⁵. Such a flowering of life arising so soon after it was physically possible to do so bears directly on the ques-

tion of the difficulty or inevitability of life beginning at all⁶.

Although the fossil date seems rock-solid, Doolittle and colleagues' contradictory evidence is hard to dismiss outright. The usual procedure in molecular analyses compares the different versions of a protein found in different organisms. The number of differences between the amino-acid sequences of each pair of species gives an 'evolutionary distance' between them, and a matrix of such distances is used to assign branch points on a phylogenetic tree. Because rates of changes may vary, different proteins may give very different trees. Doolittle *et al.* therefore took the unprecedented step of pooling data from genes for 57 different metabolic enzymes, using a total of 531 sequences representing 15 major phylogenetic groups. The trees built with the resulting evolutionary distances^{7,8} allowed fast- and slowly-evolving groups to be identified, and appropriate corrections to be made. The data are internally consistent, in that different proteins tend to give the same tree, and the pattern of branching agrees with that estimated with other molecules and other techniques^{9–11} (a in the figure).

The crucial (and controversial) step was calibrating the tree. Doolittle *et al.* began by plotting the dates when seven key animal lineages first appeared in the fossil record against adjusted sequence divergence at the seven corresponding branchpoints. For example, the earliest fossil echinoderms and chordates both date from the mid-Cambrian¹², which sets the latest possible date for the split between them at 550 million years. A straight line was then fitted to these points and extrapolated back another 2 billion



Timeline of life's history, illustrating the discrepancy between the microfossil evidence and the latest molecular analysis¹. a, Comparison of protein sequences gives a phylogenetic tree, according to which the last common ancestor lived about 1.8 billion years (Gyr) ago. b, Well-established microfossils point to a flowering of life, including modern-looking cyanobacteria, 3.5 Gyr ago.

years, allowing the dates associated with the more ancient divergences to be estimated (*a* in the figure). The discordance between these divergence dates and the early fossil record (*b* in the figure) is obvious. To restore our previous view of early evolution we must either discard the Archaean fossil evidence or move the first divergence of the Doolittle tree back almost 2 billion years. Could either of these be justified?

Neither the age nor the biological origin of the 3.5-Gyr-old microfossils has recently been seriously questioned, and the similarities between modern and fossil forms supports their common biological origin. Furthermore, the occurrence of stromatolites of different ages in the Archaean and early Proterozoic is consistent with a continuous flow of life from its inception 3.5 Gyr ago to the present. Still, the stromatolites are not unquestionably considered to be of biological origin⁵, suggesting we might do well to reconsider the oldest microfossils.

On the other hand, the molecular data may be misleading. Although the extrapolation seems unreliable (a line extended back to more than three times its original length), the discrepancy with the fossil data is so extreme that only drastic measures could reconcile the two. It could be that the reference fossils used by Doolittle *et al.* are misdated, being older than they seem. However, we should have as much or more faith in these more recent fossils as in the ancient microfossils. Perhaps we are simply missing much earlier fossils for each of the groups used in the calibration. Or, it might be that the observed rate of amino-acid substitution, which seems to be fairly constant over the time span used for calibration ($r = 0.94$ for $n = 7$ points) actually slows down far into the past, due to a slower rate of amino-acid substitution during early times. This idea has no basis in theory.

Alternatively, the model of substitution⁸ used by Doolittle and colleagues might not be sophisticated enough to adjust fully for several substitutions at the same site, and so is underestimating divergence times. Doolittle *et al.* tested for one form of departure from the standard model, keeping certain amino-acids constant, and found that this had little effect. But there is some evidence^{13,14} that the positions within proteins that cannot support a substitution may change through time, as other amino-acids around them change, and this is a much harder bias to contend with. Although the effect would be most serious for very divergent lineages, evidence, in the form of too-recent extrapolated dates, should be present throughout the tree. Without fossils, this is hard to test, but there is no glaring evidence for such a bias. Indeed, the estimated split between plants and (animals + fungi) at 1.0 Gyr may strike some as too old.

What if both dates are correct? Per-

haps, given the measurement errors associated with both fossils and molecules, we should be happy with a factor-of-two discrepancy. Or, if there was a diverse assemblage of cyanobacteria-like bacteria around 3.5 Gyr ago, and all living organisms looked at by Doolittle *et al.* have a common ancestor at 1.8 Gyr, then we could well be but footnotes to some particular cyanobacterial lineage. Or, if the resemblance between fossil and modern cyanobacterial forms is only convergent, perhaps we are descended from some other, unnamed member of the early biota. Might the original cyanobacteria-like lineage (or another branch from this time) persist? If so, it should root in the tree well below the rest of us.

In any case, we must account for the extinction of all the other lineages found in the oldest rocks. If this seems far-fetched, then consider that, if the resemblance between the fossil and modern cyanobacteria is truly convergent, it is also possible that we are the descendants of a completely different diversification. The fossil record between the early Archaean and the early Proterozoic (around 2 Gyr ago) is very sparse indeed¹. If life arose and diversified with alarming speed once, three-and-a-half billion years ago, why not again a billion-and-a-half years later? Unfortunately, this hypothesis would be hard to test.

The new study¹ is the first broadly based molecular sally into the debate on the timing of life's origin. But it is surely not the final word. Analyses of molecular evolution continue to improve, and fossils continue to be both found and re-evaluated. Only time will tell who is right — only, however, if time can. □

Arne Ø. Mooers is in the Department of Zoology, and Rosemary J. Redfield is in the Department of Zoology and the Canadian Institute for Advanced Research, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

1. Doolittle, R. F., Feng, D.-F., Tsang, S., Cho, G. & Little, E. *Science* **271**, 470–477 (1996).
2. Schopf, J. W. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6735–6742 (1994).
3. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miya, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9355–9359 (1989).
4. Schopf, J. W. & Walter, M. R. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 214–238 (Princeton Univ. Press, 1983).
5. Schopf, J. W. *Science* **260**, 640–646 (1993).
6. Thaxton, C. B. *The Mystery of Life's Origin: Reassessing Current Theories* (Lewis & Stanley, Dallas, 1992).
7. Feng, D.-F. & Doolittle, R. D. *J. molec. Evol.* **25**, 351–360 (1987).
8. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure*, Suppl. 3 (ed. Dayhoff, M. O.) 345–352 (National Biomedical Research Foundation, Washington, D.C., 1978).
9. Knoll, A. H. *Science* **256**, 622–627 (1992).
10. Wainwright, P. O. *et al. Science* **260**, 340–342 (1993).
11. Jukes, T. H. *Space Life Sci.* **1**, 469–490 (1969).
12. Benton, M. J. (ed.) *The Fossil Record 2* (Chapman & Hall, London, 1993).
13. Fitch, W. M. & Markowitz, E. *Biochem. Genet.* **4**, 579–593 (1970).
14. Fitch, W. M. & Ayala, F. J. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6802–6807 (1994).

DAEDALUS

Wind and sand

THE fluidized bed is one of the neatest of inventions. Blow a controlled current of air through a mass of fine particles, and it expands into a turbulent dense fluid. Daedalus now proposes a new use for the principle. He is inventing the fluidizing hovercraft.

Imagine, he says, a sort of raft on a shallow bed of (say) sand; and imagine air being blown down through holes in the raft to fluidize the sand beneath. The result would be a pressurized, turbulent, sand-and-air cushion under the raft. Quite a small air flow would maintain the fluidization — air would be hindered from escaping round the edges of the raft by an unfluidized periphery of compact sand. Even if the cushion exerted a mere fraction of an atmosphere of pressure, this would allow a raft ten metres square to carry several hundred tons. The sand layer need not even be very deep. Fluidization would only have to extend a centimetre or so below the underside of the raft.

The resulting load-bearing platform would be readily mobile. Push it along, and it would fluidize the fresh sand it encountered, while the sand it left behind would pack down again. It might be best propelled by wheels or tracks reaching down through the shallow fluidized cushion to the solid surface beneath.

This splendid new vehicle will find many uses. For a start, it will transform desert life. Fluidizing hovercraft will range freely over the world's sandy deserts. They could climb gradients, and could scale a dune or explore a wadi with wonderful speed and ease. They might even make progress over fairly recent snow, by the snow-blower effect, though old snow would probably be too coherent. In arable or urban regions the new hovercraft will need a special sandy track to run on. With its huge potential load-carrying capacity, it should rival the freight railway as a transporter of bulk materials such as rock, coal and iron ore. But its most revolutionary application would be to buildings, especially those on the well-known 'concrete raft' type of foundation.

Such buildings, fitted with ground-floor air pumps, could be assembled in a specialized 'factory' on the site. As each was completed, its fluidizing pumps would be turned on. It could be towed along a sandy track to its chosen position on the estate. When the pumps were turned off, it would bed down stably on the sand. But unlike a conventional building, it could always be moved again. As the estate grew or changed function, it could be reoriented, resited or towed away for modification or demolition.

David Jones

Yeast genes and human disease

SIR—The experimental advantages of the budding yeast *Saccharomyces cerevisiae* are increasingly being exploited to elucidate the function of genes mutated in human disease states. Classical and recombinant genetics approaches have, for example, allowed researchers to isolate second-site revertants for the most common mutation in the human *CFTR* gene associated with cystic fibrosis¹. The simple sequence repeat instability observed in yeast *msh2* mutants, resembling microsatellite instability in human *rer*⁺ tumours, led researchers to clone the human homologue *hMSH2*, mutated in one form of hereditary non-polyposis colon cancer^{2,3}. The study of yeast *MEC1* and *TEL1* genes has offered insights into the function of the human *ATM* gene, mutated in ataxia telangiectasia^{4,5}, and provided an experimental system for further analysis.

To date, 70% of the *S. cerevisiae* genome has been deposited in public databases, and the entire sequence is expected early this year. This will add to the already powerful repertoire of experimental advantages in this model organism (see the article by Oliver on page 597 of this issue⁶). The genome sequence will

also allow cross-species connections that directly benefit human disease research to be made with greater frequency; but how often can we expect these connections to be made, and how can we increase the rate at which they are established?

To address the first question, we have compiled a table of sequence similarity between all positionally cloned human genes (48 to date) and *S. cerevisiae* genes. By virtue of their method of identification, the set of positionally cloned genes is relatively unbiased in tissue specificity and level of expression. Many of these genes (25% at present) match an *S. cerevisiae* protein at least as significantly as the human *NF1*-yeast *IRA2* pair (P value = 2.0×10^{-46} ; see table) — a reasonable benchmark, as the human *NF1* complementary DNA expressed in yeast can complement a yeast *ira2* mutant⁷. This percentage will almost certainly increase as the remaining 30% of the yeast genomic sequence becomes available in the coming months. The complete table, containing all positionally cloned human genes, is available on the World Wide Web, along with a version which will be kept updated (see table legend).

The relationship between model organism genes and genes mutated in human diseases can also be used to discover new disease genes. The EST Division of GenBank (dbEST)⁸ has already been used successfully to identify human homologues of genes previously cloned in model organisms. In assessing the current potential of this approach, we have determined that, at present, 79% (38/48) of the positionally cloned human genes are represented by exact DNA sequence matches with one or more ESTs in dbEST, up from 38% only one year ago (see http://www.ncbi.nlm.nih.gov/dbEST/dbEST_genes/). A significant portion of the human gene repertoire is thus already likely to be represented in dbEST, and this figure will improve as dbEST grows.

Finally, a comparative genomics database, XREFdb, has recently been developed for public use. This free resource is dedicated to establishing cross-references between human genes and biological information available for homologous genes in model organisms, and expediting the identification of genes mutated in human disease using an EST-based approach⁹. XREFdb provides each user with automatic monthly BLAST similarity searches and tracking of protein and EST matches for queries of interest, and contains mouse and

BEST SEQUENCE SIMILARITY MATCHES TO DATE BETWEEN POSITIONALLY CLONED HUMAN GENES AND *S. CEREVISIAE* PROTEINS

Human disease	MIM no.	Human gene	GenBank acc. no. for human cDNA	BLASTX <i>P</i> -value	Yeast gene	GenBank acc. no. for yeast cDNA	Yeast gene description
Hereditary non-polyposis colon cancer	120436	<i>MSH2</i>	U03911	9.2×10^{-261}	<i>MSH2</i>	M84170	DNA repair protein
Hereditary non-polyposis colon cancer	120436	<i>MLH1</i>	U07418	6.3×10^{-196}	<i>MLH1</i>	U07187	DNA repair protein
Cystic fibrosis	219700	<i>CFTR</i>	M28668	1.3×10^{-167}	<i>YCF1</i>	L35237	Metal resistance protein
Wilson's disease	277900	<i>WND</i>	U11700	5.9×10^{-161}	<i>CCC2</i>	L36317	Probable copper transporter
Glycerol kinase deficiency	307030	<i>GK</i>	L13943	1.8×10^{-129}	<i>GUT1</i>	X69049	Glycerol kinase
Bloom's syndrome	210900	<i>BLM</i>	U39817	2.6×10^{-119}	<i>SGS1</i>	U22341	Helicase
Adrenoleukodystrophy, X-linked	300100	<i>ALD</i>	Z21876	3.4×10^{-107}	<i>PAL1</i>	L38491	Peroxisomal ABC transporter
Ataxia telangiectasia	208900	<i>ATM</i>	U26455	2.8×10^{-90}	<i>TEL1</i>	U31331	PI3 kinase
Amyotrophic lateral sclerosis	105400	<i>SOD1</i>	K00065	2.0×10^{-58}	<i>SOD1</i>	J03279	Superoxide dismutase
Myotonic dystrophy	160900	<i>DM</i>	L19268	5.4×10^{-53}	<i>YPK1</i>	M21307	Serine/threonine protein kinase
Lowe's syndrome	309000	<i>OCRL</i>	M88162	1.2×10^{-47}	<i>YIL002C</i>	Z47047	Putative IPP-5-phosphatase
Neurofibromatosis, type 1	162200	<i>NF1</i>	M89914	2.0×10^{-46}	<i>IRA2</i>	M33779	Inhibitory regulator protein

As of 20 January 1996, 25% (12/48) of positionally cloned genes mutated in human disease states match an *S. cerevisiae* gene in GenBank at least as significantly as the human *NF1*-yeast *IRA2* pair ($P = 2.0 \times 10^{-46}$). BLASTX searches were performed using each positionally cloned human cDNA listed against a non-redundant database of *S. cerevisiae* sequences (NRSC, 20 January 1996) maintained by the Saccharomyces Genome Database project (<http://genome-www.stanford.edu/>). The BLASTX (v1.4) *P*-value for the most statistically significant yeast protein match for each human query is given, along with the GenBank accession number for the corresponding *S. cerevisiae* nucleotide sequence (ordered by increasing *P*-value). Where a GenBank entry for the complete cDNA sequence for a positionally cloned gene does not exist, the GenBank accession number for the longest available partial sequence is listed. BLASTX searches were performed using default parameters. The SEG algorithm was used for all searches to mask low-complexity sub-sequences. Further information for each human disease is available through the Online Mendelian Inheritance in Man (OMIM) database (<http://www3.ncbi.nlm.nih.gov/Omim/>) using the MIM numbers provided. A more complete version of the table is available online through the World Wide Web which will be kept up to date as the *S. cerevisiae* genome sequencing project progresses, and as more human genes are positionally cloned (see <http://www.ncbi.nlm.nih.gov/Bassett/Yeast/>).

human EST mapping data as well as human phenotype data (to access XREFdb, see <http://www.ncbi.nlm.nih.gov/XREFdb/>).

Douglas E. Bassett Jr^{††}

Mark S. Boguski[†]

Philip Hieter^{*}

^{*}The National Center for Biotechnology Information,

National Library of Medicine,
National Institutes of Health,
Bethesda, Maryland 20894, USA

[†]Department of Molecular Biology and Genetics,

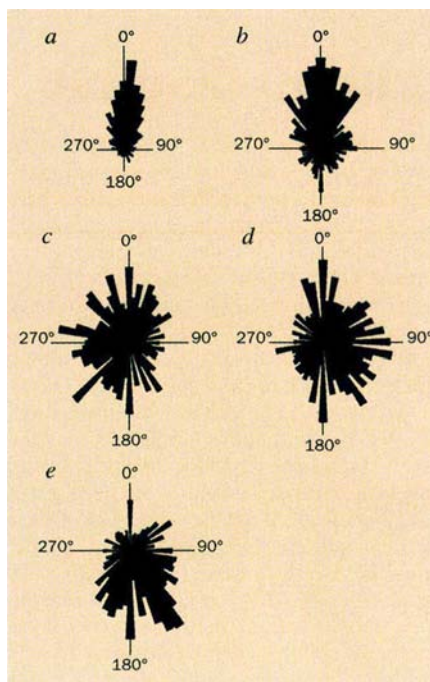
The Johns Hopkins University
School of Medicine,
Baltimore, Maryland 21205, USA

1. Teem, J. L. *et al. Cell* **73**, 335–346 (1993).
2. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. *Nature* **365**, 274–276 (1993).
3. Fishel, R. *et al. Cell* **75**, 1027–1038 (1993).
4. Weinert, T. A., Kiser, G. L., Hartwell, L. H. *Genes Dev.* **8**, 652–665 (1994).
5. Morrow, D. M., Tagle, D. A., Shiloh, Y., Collins, F. S. & Hieter, P. *Cell* **82**, 831–840 (1995).
6. Oliver, S. G. *Nature* **379**, 597–600 (1996).
7. Ballester, R. *et al. Cell* **63**, 851–859 (1990).
8. Boguski, M. S., Tolstoshev, C. M. & Bassett, D. E. Jr *Science* **265**, 1993–1994 (1994).
9. Bassett, D. E. Jr *et al. Trends Genet.* **11**, 372–373 (1995).

How *Euglena* tells up from down

SIR—The green, unicellular flagellate *Euglena gracilis* orientates itself in the water column by means of external stimuli such as light and gravity^{1,2}. Although much is known about photo-orientation, we know little about gravity-dependent orientation of microorganisms. Here we show that gravitactic orientation is strongly impaired by an increase in the density of the surrounding medium, by inhibitors of stretch-sensitive ion channels and by the disturbance of membrane potentials.

The density of the medium was increased using Ficoll (see figure). At low Ficoll concentrations (2.5%), *Euglena*'s negative gravitaxis (orientation away from the centre of gravity) was only slightly impaired (*a, b* in the figure). With increasing concentrations, the density of the medium approached the density of the *Euglena* cell body (1.04 g ml⁻¹) and negative gravitactic orientation was strongly impaired (*c, d* in the figure). At higher concentrations, the cells orientated towards the centre of gravity (*e* in the figure).



Circular histograms of gravitactic orientation of *E. gracilis* cells in different Ficoll concentrations: *a*, control; *b*, 2.5%; *c*, 5%; *d*, 7.5%; and *e*, 10%. 0° indicates upwards. Cells were incubated for 4 h in the dark to enhance gravitaxis. The speed of movement and orientation of organisms were determined using a real-time, computer-based cell-tracking system⁷. A typical experiment analysed about 1,000 cell tracks.

TPMP⁺, a lipophilic cation, passes through membranes following the existing membrane potential and thus tends to nullify the potential. Even at low concentrations (10 μM), TPMP⁺ strongly impaired gravi-orientation (data not shown), showing that the membrane potential is crucial for gravitactic orientation. Stretch-sensitive ion channels can be inhibited by gadolinium ions (Gd³⁺)^{3–5}. Low concentrations of Gd³⁺ (100 μM) strongly affected gravi-orientation. Gravitaxis was totally inhibited after 90 s (data not shown), strongly indicating the existence and involvement in gravitaxis of stretch-sensitive ion channels in the cell membrane of *E. gracilis*.

Based on the finding of a density difference between *Euglena*'s cell body and the medium, we propose a mechanism for gravi-orientation using a hypothesis already introduced for gravitaxis in ciliates⁶. The resulting pressure of the cell body deforms different parts of the cell membrane, depending on the position of the cell. These deformations activate stretch-sensitive ion channels, and the resulting change in the ion flux alters the membrane potential and is the primary event of gravity-sensing. In addition, we predict that stretch-sensitive ion channels involved in gravitaxis are not evenly distributed in the cell membrane. Channels are probably concentrated at either the

poles or the sides of the cell, depending on the mode of activation. In a vertical swimming direction, the modulation of the membrane potential caused by the ion channels would be minimal.

The hypothesis that the general body surface can give unicellular organisms, which are heavier than the surrounding medium, a means of determining their orientation with respect to gravity might be used by numerous protozoa showing orientation responses.

Michael Lebert

Donat-Peter Häder

Institute for Botany and Pharmaceutical Biology,
D-91058 Erlangen, Germany

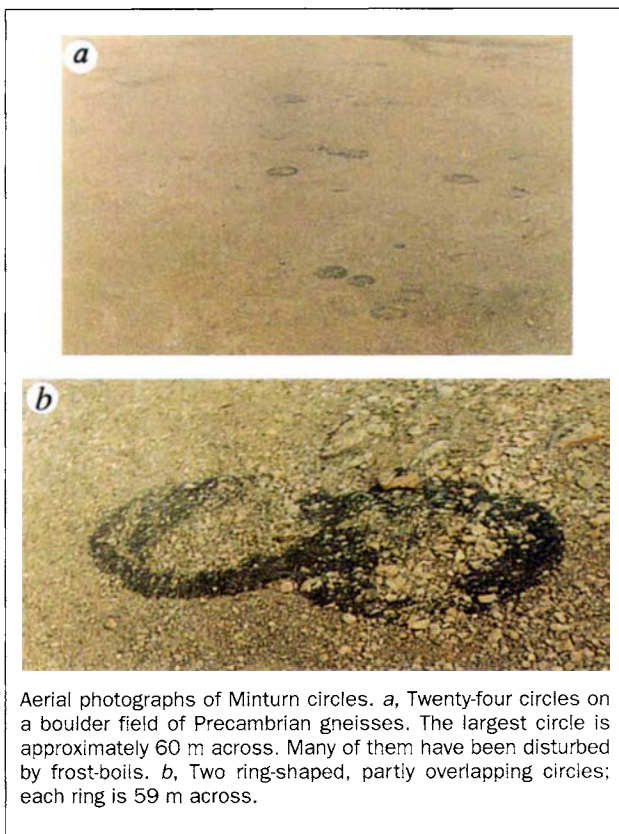
A new type of glacial deposit

SIR—In Inglefield Land, northwest Greenland, a new type of glacial deposit has been discovered on the flat plains between the Inland Ice and Kane Basin. These circular and ring-shaped deposits are named Minturn circles after the Minturn Elv, a turbulent river traversing Inglefield Land.

The Minturn circles are well-defined, occasionally overlapping rings and circles (see figure) of various colours (mostly black to grey). Around 300 were found in a fan-shaped belt perpendicular to the margin of the Inland Ice. The widths of the rings vary from less than 1 metre to 8 m, and have a sharp outer periphery, although the inner rim of the rings may be diffuse. Rings vary in diameter from less than 1 metre to 8 m, while circles tend to be larger. They are found mainly on flat surfaces such as boulder fields and flat bedrock, but several are seen on slopes of up to 45°. These features are composed of an unsorted assemblage of slightly rounded boulders, cobbles and pebbles of syenite, which form a layer varying in thickness from a few centimetres to 3 m. Mostly it is a thin sheet one block thick, covering a subsurface of gneisses and sediments. The syenite blocks are often covered by black lichen (the feature that led to their discovery), whereas the less fertile surrounding gneisses and sediments support less lichen.

The first stage in formation of the Minturn circles seems to have occurred under the Inland Ice in an area where the ice was periodically frozen to the ground. There, a melting–refreezing process took place, similar to that described elsewhere^{1,2}, whereby debris-rich ice-sheets were formed and gradually transported across Inglefield Land by the Inland Ice. The composition of the blocks found in the Minturn circles indicates that the melting–refreezing process took place above a large syenite complex. During

1. Häder, D.-P. *Arch. Microbiol.* **147**, 179–183 (1987).
2. Häder, D.-P., Rosum, A., Schäfer, J. & Hemmersbach, R. *J. Pl. Physiol.* **146**, 474–480 (1995).
3. Franco, A. Jr, Winegar, B. D. & Lansman, J. B. *Biochem. J.* **59**, 1164–1170 (1991).
4. Lacampagne, A., Gannier, F., Argibay, J., Garnier, D. & Le Guennec, J.-Y. *Biochim. biophys. Acta* **1191**, 205–208 (1994).
5. Sukharev, S. I., B. Martinac, V.Y. Arshavsky & Kung, C. *Biophys. J.* **65**, 177–183 (1993).
6. Machemer, H. & Bräucker, R. *Acta protozool.* **31**, 185–214 (1992).
7. Häder, D.-P. *Binary* **6**, 81–86 (1994).



Aerial photographs of Minturn circles. *a*, Twenty-four circles on a boulder field of Precambrian gneisses. The largest circle is approximately 60 m across. Many of them have been disturbed by frost-boils. *b*, Two ring-shaped, partly overlapping circles; each ring is 59 m across.

melting of the Inland Ice, the syenite blocks became exposed on the flat surface of the ice.

Transport of melt water from a cold ice cap occurs at the surface in a multitude of meandering streams which may evolve around segments that are gradually isolated from the active streams as more or less

ring-shaped structures. Superficial moraine will slide down into these isolated ring-shaped structures. In some cases, thin moraine may be deposited, whereas in others moraine could fill the ring structures. In the latter case, solar heating of the moraine may cause subsequent melting whereby the outer and inner rim melts away and a circle is produced.

One problem with this explanation is that cut-off segments from meandering streams rarely attain a perfectly round shape. Alternatively, Minturn circles may have formed in pot-holes made by streams on the surface of the Inland Ice. When the Inland Ice finally melted, the circular and ring-shaped accumulations of syenite debris would then have been

deposited very gently.

Peter W. Uitterdijk Appel
Geological Survey of Denmark
and Greenland,
Øster Voldgade 10,
DK 1350 Copenhagen, Denmark

1. Weertman, J. *J. Glaciol.* **3**, 965–978 (1971).
2. Boulton, G. S. *J. Glaciol.* **9**, 213–229 (1970).

A newly defined interleukin-1?

SIR — Okamura *et al.*¹ report the cloning of a novel, stress-induced cytokine from mouse liver which has biological actions in T cells that are strikingly reminiscent of interleukin-12 (IL-12). The polypeptide chain of this molecule, termed interferon- γ -inducing factor, or IGIF, has no apparent similarity to sequences in protein or nucleotide databases¹. To pursue further the possible kinship of IGIF to existing

cytokine families² (notably the haematopoietic group of regulators that includes IL-12), we have used fold recognition methods to classify IGIF, as protein structure is often a better measure of evolutionary relationship than is primary sequence.

Two algorithms that rely on very different potential functions to score sequence–structure compatibility —

123D (ref. 3) and ProFIT (ref. 4) — threaded the mature IGIF sequence through protein fold databases and drew consistent, top matches to the IL-1 β structure (represented by Protein Data Bank locus 411B), a fold unrelated to haematopoietic factors². A structurally based comparison of IGIF with IL-1 α , IL-1 β and receptor antagonist (IL-1ra) chains (see figure) shows a parsimonious alignment over the length of the respective sequences that comprise the twelve β -strands of the distinctive ‘ β -trefoil’ fold⁵. Significant stretches of similarity indicated by this alignment are further evidence of the IGIF homology to the IL-1 family. A corresponding three-dimensional IGIF model built from the IL-1 β framework displays the robust features of a native-like protein fold^{4,6}.

This structural proposition has several important implications. The ‘unusual leader sequence’ of IGIF¹ may be analogous to the IL-1 β pro-domain, which must be cleaved for optimal secretion and biological activity; IGIF maturation at the Asp¹Asn site is probably mediated by an IL-1 β convertase-like enzyme⁷. Also, an IL-1 binding paradigm, based on extensive mutagenesis studies⁷ and recently defined by the receptor complex structure of IL-1ra (ref. 8), may extend to the recognition of IGIF by a signalling complex that perhaps involves IL-1-receptor subunits. If this model holds, IGIF could exert its biological actions with an IL-1-like NF- κ B signalling strategy⁷, distinct from the JAK/STAT pathway of IL-12.

We suggest that IGIF is a member of the IL-1 family of cytokines that function in the immune and inflammatory response⁷, and tentatively propose a structurally accurate designation of ‘IL-1 γ ’.

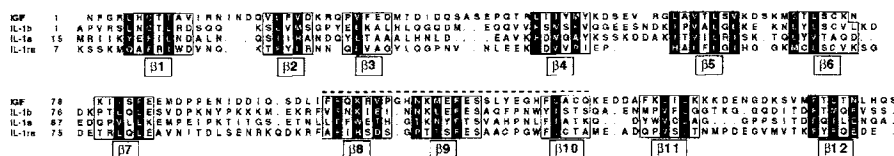
J. Fernando Bazan

Jackie C. Timans

Robert A. Kastelein

Department of Molecular Biology,
DNAX Research Institute,
Palo Alto, California 94304, USA

1. Okamura, H. *et al. Nature* **378**, 88–91 (1995).
2. Sprang, S. R. & Bazan, J. F. *Curr. Opin. Struct. Biol.* **3**, 815–827 (1993).
3. Alexandrov, N. N., Nussinov, R. & Zimmer, R. M. *Pacif. Symp. Biocomputing '96* (eds Hunter, L. & Klein, T. E.) 53–72 (World Scientific, Singapore, 1995).
4. Flockner, H., Braxenthaler, M. & Sippl, M. *J. Proteins* **23**, 376–386 (1995).
5. Murzin, A. G., Lesk, A. M. & Chothia, C. *J. molec. Biol.* **223**, 531–543 (1992).
6. Levitt, M. *J. molec. Biol.* **226**, 507–533 (1992).
7. Dinarello, C. A. *FASEB J.* **8**, 1314–1325 (1994).
8. Schreuder, H. *Cytokine* **7**, 599, Abstr. A23 (1995).



Structural alignment of IGIF with human IL-1 cytokines. Representative IL-1 α , IL-1 β and IL-1ra folds are superposed following Murzin *et al.*⁵ and aligned to the IGIF sequence by threading^{3,4}, structure prediction and conserved residue patterns⁵. β -Strands are boxed and numbered 1 to 12; white residues contribute to the hydrophobic core⁵; the dashed bar marks the region of highest sequence similarity; dots designate alignment gaps; italicized cysteines in IL-1ra form the only disulphide bridge in this family. IGIF exhibits 19% and 12% positional identity to IL-1 β and IL-1 α , respectively; by comparison, the latter two molecules show 23% sequence identity.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

An Early Silurian vascular plant

SIR—What the earliest vascular plants looked like, and when and where they began to appear on Earth, are questions that have concerned and interested many people. *Cooksonia*, originally recorded from the Late Silurian of England, is generally accepted as the first vascular plant. The co-occurrence of vascular plant remains, including *Pinnatiramosus* and single-retusoid trilete spores in the Early

Silurian of southwest China, however, demonstrates that they must have had an earlier history. Remains of *Pinnatiramosus*, first discovered by the 108 Regional Geological Team of Guizhou in 1972, were formally described and named *Pinnatiramosus qianensis* by Geng¹, who suggested that it represented a stock of extinct plants with algal external morphology and tracheo-phyte anatomical features, but closer relationship with some algae. Our studies on a new sample of this plant² indicate that it may be the earliest-known vascular plant.

This plant is found in an outcrop section at Dongkala of Fenggang County, Guizhou, southwest China, in strata considered Wenlock in age¹. We conclude that they may correspond to Late Llandovery, as evidenced by associated brachiopods, chitinozoans, bivalves (Fig. 1) and regional stratigraphic correlation³. The main features of *Pinnatiramosus* include prostrate axes and pinnate branches interpreted as erect by Geng (Fig. 2). Structures are found in the prostrate axis which resemble tracheids with bordered pits. Similar structures are present in associated plant debris. This kind of tracheid has never been recorded in lower plants (algae and bryophytes) and is characteristic of vascular plants.

The palynomorphical assemblage is dominated numerically by obligate tetrads (around 70% of the total) in association with diverse single-trilete spores, including six to eight species of *Punctatisporites*, *Ambitisporites*, *Retusotriletes* and *Apiculiretusispora*. Most of these single spores are retusoid and characteristic of ancient vascular plants. Tetrads are represented by *Nodosphaera* and *Tetrahedraletes*, and acritarchs by *Leiosphaeridia* and *Visbyphaera*. The assemblage may be equivalent to the *avitus-dilutus* zone or the base of the *chulus* var. *chulus-chulus* var. *nanus*

zone of Richardson and McGregor⁴, or near the base of zone II of Gray⁵.

The debate about *Pinnatiramosus* concerns its strange external morphology and complex internal anatomy, which is barely comparable to any known vascular plants from the Silurian and Early Devonian. Some authors suspected that the plant remains were derived from the overlying younger strata⁶; some assume that these root-like branches might penetrate downward from the overlying Permian into the Silurian strata, and the spore data seem not to fit with ages based on invertebrates, as they show some aspects of Wenlock-Ludlow assemblages.

We argue that: (1) the plant-bed is 2.2 m below the base of the Permian strata and no fossil plants, including rootlets, have been found in-between; (2) the plant remains are mainly preserved in fresh, solid and obviously unweathered, yellowish green siltstone, not suitable for root penetration by primitive vascular plants; (3) the fossil plants are indigenous and mostly preserved parallel or obliquely to bedding planes, and seem to have been buried *in situ* or undergone only short movement by water currents; (4) the tracheid-like structures are within the megafossil and the tracheids closely resemble those of modern tracheotypes, at least in outer morphology; (5) the palynological data indicate an age younger than Aeronian and probably older than Late Wenlock. The dominance of obligate tetrads seems to support the dating by invertebrates, as the change from spore tetrads to single-trilete spores has been placed in the mid-to-late Early Silurian⁵. Palynological data from the Llandovery are very rare, and so any generalization has only limited significance.

From the above evidence and discussion, we conclude that *Pinnatiramosus* is indeed a Silurian plant, rather than root-like branches derived from Permian, and is most likely to be a vascular plant. The plant-bearing bed is Late Llandovery in age and the spore data also show that vascular plants existed by the Early Silurian.

Chongyang Cai

Shu Ouyang

Yi Wang

Zongjie Fang

Jiayu Rong

Liangyu Geng

Xingxue Li

*Nanjing Institute of Geology & Palaeontology,
Academia Sinica,
Nanjing 210008, China*

1. Geng, B. *Acta bot. sin.* **28**, 664–670 (1986).

2. Cai, C. & Li, X. *Palaeont. cath.* **6**, 167–214 (1994).

3. Rong, J. et al. *J. Stratigr.* **14**, 161–177 (1990).

4. Richardson, J. B. & McGregor, D. C. *Bull. geol. Surv. Can.* **364**, 1–79 (1986).

5. Gray, J. *Phil. Trans. R. Soc. B309*, 167–195 (1985).

6. Edwards, D. *Geol. Soc. Mem.* **12**, 233–242 (1990).

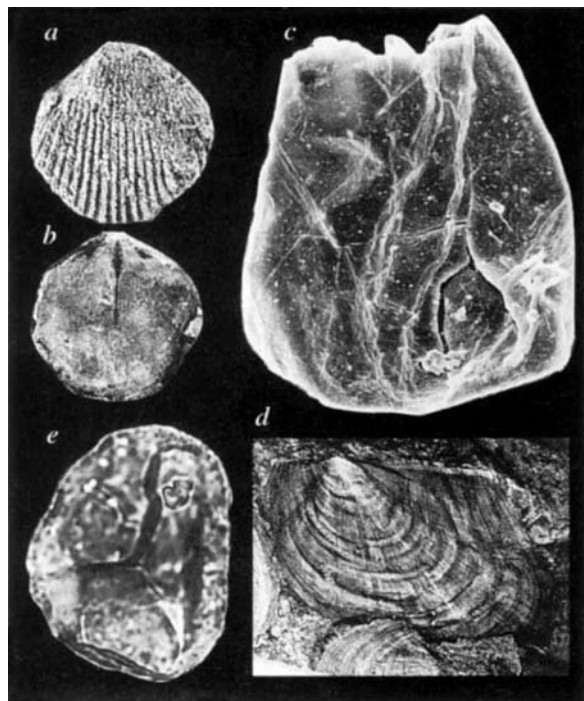


FIG. 1 Fossils associated with *P. qianensis*. a, *Nalivkinia* sp. ($\times 1.78$); b, *Nucleospira pulchra* ($\times 1.78$); c, *Eisenackitina daozenensis* ($\times 445$); d, *Pterinopecten* cf. *cybele* ($\times 1.78$); e, *Apiculiretusispora* sp. ($\times 712$).

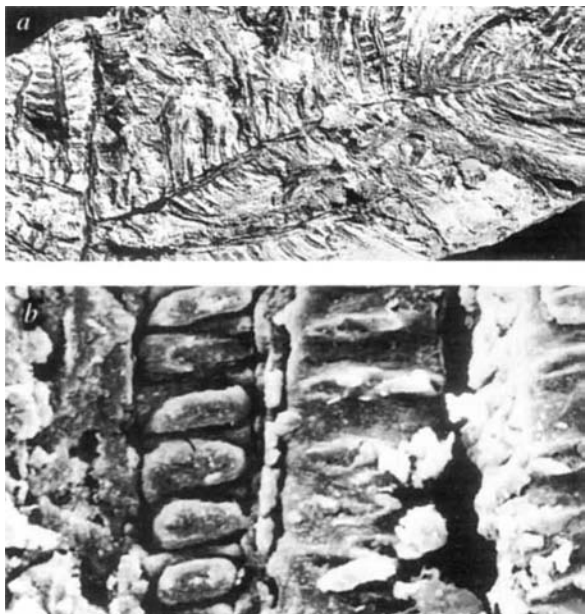


FIG. 2 a, External morphology of *P. qianensis*; b, scanning electron micrograph of the tracheids with bordered pits *in situ*.

Book of numbers

Vaclav Smil

How Many People Can the Earth Support? By Joel E. Cohen. Norton: 1995. Pp. 532. \$30, £22.50. UK publication date, 27 March.

THE reader will never get the definitive answer to the question posed by the title of this book. That is, of course, the way it should be. As the author knows well, any attempt to offer and justify such a number is misplaced; the question has no single answer today and it will not have one tomorrow — at least, none from unprejudiced enquiry. Naturally, the catastrophists who long ago decided that the world already has too many people and the techno-fixers who are itching to terraform Mars all have their own answers: no more than about a billion in the first case, an irrelevant astronomical total in the second.

Those of us who prefer reality to dogma know that all sorts of natural, technical and social factors constrain human choices about the planet's carrying capacity, but that our understanding of these limits constantly evolves and changes, and that their acuteness and permanence are vigorously debated, disputed and, in some instances, even denied. If one insists on numbers, then the only sensible way out is to give conditional estimates: that is, given such and such a combination of future choices, available resources, techniques and social arrangements, then it is most likely that the Earth could support so many people. Usually it is only a short time before the many requisite assumptions become either completely indefensible or highly questionable, and one may then begin a new round of scenario-building. These are futile exercises, but the question is here to stay.

One way for a writer to dispose of the question would be to flip it aside by pointing out some of these key realities in a snappy essay. But Joel Cohen chooses instead to smother it. The essay is still there — but the reader must first wade through some 260 pages to reach it. The main reason for the postponed appearance is clear. The author, head of the Laboratory of Populations at Rockefeller University, New York, has done a great deal of interdisciplinary work, but in this book, which offers a perfect opportunity for including the widest range of disciplines and perspectives, he is very much a demographer.

Before attempting to answer the unanswerable, he devotes a lot of space (too much, I think) to a systematic description of past population growth and its long-

term projections. He takes the first topic, which consumes a quarter of the text, from uncertain estimates of total human numbers at the time of the emergence of agriculture to an extended discussion of various growth curves. Forecast reviews, occupying the next quarter of the book, look first at various projection methods and scenarios and then in extensive detail — with summaries of original publications followed by the author's comments — at eight different estimates of human carrying capacity published, by individuals or teams, between 1891 and 1991.

Only then does Cohen start moving beyond the demographic confines by looking at the concept of carrying capacity in ecological terms and by assessing its worth for humans. Again, this is done in a very deliberate way, including an appendix of two dozen verbal definitions of human carrying capacity printed in small type.

None of this is unimportant, most of it is relevant and much of it is interesting. But who is it aimed at? Norton books are

intended to reach a wide readership, but the material here may simply be too much for a lay reader and is mostly too well known to be of value to any serious student of human population. I believe the dilatory stratagem works against the book. Entering *in medias res*, after just a crisp brief recount of the past — which, in this case, is no guide for the future — would have been much better.

Two-thirds into the book, we come to its real core, to what it is all about: human choices. To me the single most important part of the book is a list of 11 simple questions, beginning with "How many at what average level of material well-being?" and "How many with what distribution of material well-being?" and ending with "How many for how long?" and "How many with what values, tastes and fashion?". Reflecting on these questions and trying to answer them in both a personal and a detached way would make any serious reader richly aware of the almost exhilarating indeterminacy of the outcome — and of the immense individual and collective responsibility for the fate of the species.

Unfortunately, Cohen then soldiers on for another hundred pages. Realizing the enormous variety of constraints and opportunities that will shape the human future, he chooses a single case study to illustrate the range of long-term predictions of carrying capacity. He looks at



BETWEEN 1917 and 1920, in the largest internal migration in North America's history, some 450,000 African-Americans left the rural south in search of employment and freedom in the north. The move was commemorated in a series of paintings by the American artist Jacob Lawrence entitled *The Migration of the Negro*. This panel, painted in the 1940s, shows people queuing at a railroad for tickets to the industrial cities of the north. It appears in *The Settling of North America: The Atlas of the Great Migrations into North America from the Ice Age to the Present* edited by Helen Hornbeck Tanner. The text is written by a team of acclaimed historians and illustrated with 100 full-colour maps as well as with photographs, illustrations and detailed timelines. Macmillan Publishing USA, \$39.95.

water as a natural constraint, and he does a good job in discussing its resources and requirements and its links to other growth constraints. He also presents his own calculations of upper limits on world population depending on the availability of water, on the fraction that can be used for irrigation and on prevailing diets. But the outcome is predictable: a huge fan of plausible future population totals.

Afterwards the author discusses, rather gloomily, natural constraints and time. He fears we may fail to come up soon enough with many new fundamental fixes (be they fusion energy or new staple cultivars) for easing or removing some looming constraints. This is perhaps the weakest part of the book, because it largely ignores the enormous opportunities for cumulatively immense changes that humans can produce with existing understanding and skills. Why, for example, worry whether commercial fusion will be a reality by 2040 when rational use of energy in rich countries, which consume three-quarters of all fuels and electricity, could halve the average per capita consumption within a generation, and do so while increasing the quality of life? We can make that choice, but we clearly prefer not to, although we did when world oil prices were rising. Since 1985, however, energy use in rich countries has been increasing again, even in once-so-frugal Japan.

Cohen concludes that the human population is now entering the zone at which many scholars have set its upper limit. Let us suppose they are right, and let us also assume that we will follow all recommendations the author makes in closing the book (population control, mutual aid and so on) and that our numbers will remain below some dreaded unmanageable total (supply your own particulars). Will this really make it easier to do what we cannot do today — run our economies as subsystems of the biosphere and treat our politics as an exercise in altruistic feelings? Seen in this light, moving in 'the zone' does not seem to me any more perilous than before we entered it. Could matters change were we to return to it, or well below it, after a catastrophic population collapse? Nobody knows. The only certainty is that choices will remain. □

Vaclav Smil is in the Department of Geography, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

■ **Renewable Energy Strategies for Europe: Foundations and Context** by Michael Grubb has just been published by The Royal Institute of International Affairs and Earthscan at £12.95 (pbk). It examines the political, economic and intellectual driving forces behind a renewable energy policy in Europe; projections, options and current initiatives; and the lessons that can be drawn from history.

Linguistic leaps and bounds

Paul Fletcher

Language and Human Behaviour. By Derek Bickerton. *University of Washington Press/UCL Press: 1995. Pp. 180. \$24.95, £14.95.*

It is not difficult for a linguist to feel sympathy with Derek Bickerton's assertion that modern linguistic scholarship has been neglected in discussions of human evolution. But his aims, in this book-length version of a series of lectures delivered at the University of Washington in 1992, go much further than the mere re-statement of contemporary claims and findings of linguistic theory. One of Bickerton's objectives is to enthrone linguistics in its rightful place as the prince of the behavioural sciences. If language is the means by which our species developed intelligence, rational thought and a distinctive consciousness, rather than a by-product of evolved cognitive complexity, then the specifics of syntactic theory should have broad appeal to students of human evolution. A closer look at some universal principles of grammar should make it clear that "the complexities of language are its own special complexities, arising from sources very different from the desire, conscious or otherwise, to make complex cultures manipulable".

Bickerton is aware that his thesis of the primacy of language in evolution turns contemporary discussion on its head. Language is, he claims, generally held in accounts of human evolution to be a means of communication, much more complex than that of other species, developed to deal with an increasingly complex physical and social environment. Both of these views are dismissed. Communication is just one use for language. And, far from being an invention of an increasingly clever species, language, and specifically the syntactic engine that drives it, was the evolutionary development responsible for the "enhancement of intelligence".

Radical reinterpretations, if they are to have more than the briefest half-life, demand sophistication in argument, passion in exposition and good data. Bickerton shows the first two qualities in abundance. Good data on the evolution of language is, however, hard to find. As Roger Brown observed: "there is no written record of the event of origination, and there are no fossilised phonemes or shards of ancient grammar". To compensate for the deficit, Bickerton makes use of those sources of evidence that are available and which have allowed discussion of the ori-

gins of language to become once more respectable.

Apart from the fossil record, which in terms of evidence about brain size and its implications for language is dismissed in the book, there are two other kinds of information relevant to the argument, language invention and the language abilities of nonhuman primates such as chimpanzees. Language invention — the spontaneous appearance of syntactically organized sequences *de novo* — occurs or has occurred in creoles. These are, when they first appear, novel languages created by children exposed to a variety of languages as well as a pidgin. The similarities in the structures of creoles created at different times and in different places, and the fact that they have syntactic features such as tense markers, as do all established languages, lead Bickerton to the conclusion that the creoles rely on the same biologically given neural substrate that supports children's language acquisition in more usual circumstances.

The second part of the argument is to deny the availability of this substrate to nonhuman primates. However sophisticated their abilities to demonstrate, through the production of items from a taught sign language, that they can learn a vocabulary, these species cannot organize sequences of signs syntactically. If syntax (now seen as the defining characteristic of human language) is not avail-

New in paperback

The Gene Wars: Science, Politics, and the Human Genome by Robert Cook-Deegan.

Norton, \$14.95, £10.95. "A scrupulous...narrative of the genesis and opening years of the genome projects.... An essential history", wrote Horace Freeland Judson in a review in *Nature* **371**, 753 (1994).

Flexible Bodies: The Role of Immunity in American Culture from the Days of Polio to the Age of Aids by Emily Martin.

Beacon, \$14. In this contentious blend of anthropology and cultural criticism, Martin attempts to show how immunological ideas of "flexibility" have shaped the way we think not only about our bodies as a whole but also about our economies and societies. Contains many popular representations of the immune system reproduced from magazines, comics, newspapers and books.

Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics by Lawrence Sklar.

Cambridge University Press, £13.95, \$19.95. "The book occupies itself with foundations and touches on most of the crucial issues... the only available modern text that has set itself this task", wrote Peter Landsberg in *Nature* **368**, 506 (1994).

able to our genetic near neighbours but is a "biologically based attribute of [our] species", when in evolution did it appear? In keeping with his claim that language drives cognition in the development of the species, Bickerton eschews a gradualist approach and subscribes to a syntactic "big bang" theory. In this view, the evolution of language is not seen as a gradual development from "proto-language" through intermediate stages to natural languages as we know them; rather, the transition to true language was a "catastrophic event".

This provocative thesis deserves to attract attention. Because of the potential for alternative interpretations

of the available evidence, the debate is bound to be lively. For example, although no-one is likely to dispute the central importance of syntactic mechanisms to human language, not everyone will be able to imagine as easily as Bickerton the sudden emergence of syntax in evolution or be quite as ready to draw a clear line between what he refers to as 'protolanguage' and the fully developed instances of languages in the human species. His opponents will, however, have to be on their toes to win the contest. □

Paul Fletcher is in the Department of Speech and Hearing Sciences, University of Hong Kong, 34 Hospital Road, Hong Kong.

The price of greatness

Peter Bryant

Bruno Bettelheim: The Other Side of Madness. By Nina Sutton. Duckworth: 1995. Pp. 524. £25.

BRUNO Bettelheim was a famous psychologist with an unusual background. He was born at the turn of the century in Vienna, and after a brief spell at university took over the family timber company. But he was a businessman with deep intellectual interests — particularly in psychoanalysis — which eventually drove him back to university where he completed a doctorate in aesthetics shortly before Hitler and his army marched into Austria. Partly because he was a Jew and partly because of his past political history, Bettelheim was promptly sent to Dachau and later to Buchenwald. He spent roughly a year in these dreadful concentration camps and after release moved to the United States, where he became an academic and was later put in charge of the Orthogenic School for disturbed children in Chicago — a school that had fallen on hard times but which became widely known and had some remarkable successes largely due to Bettelheim's own extraordinary work. It was there that he pioneered psychoanalytic methods for treating autistic children, techniques which have been adopted by therapists in many different countries. At the same time, he published a series of arresting and provocative books and papers on people's experiences and behaviour in concentration camps, and, later, on such diverse subjects as the effect of being brought up in an Israeli kibbutz, on the significance of fairy stories and on learning to read.

At the time of his death in 1990 he seemed to have shown an exemplary set of qualities in his work and to have thoroughly deserved the fame and the rewards that came his way. That was Bet-

telheim's reputation when Nina Sutton began her research on his life in the same year. But the picture changed abruptly a few months later when some former patients of the Orthogenic School asserted that Bettelheim had often been caustic and short-tempered with them and that quite often he resorted to violence. One of the most appalling stories of his irascibility was told by someone who claimed that Bettelheim



Enchanting dreamer: Bettelheim was a powerful but fallible therapist.

burst in on her while she was having a shower, dragged her out by the hair and then berated her in front of the other children.

These allegations were deeply shocking, not just because they were of behaviour that was quite indefensible, but also because they seemed to fly in the face of the central theme of Bettelheim's arguments about the fate of prisoners in concentration camps and also of emotionally disturbed children. In both cases Bettelheim claimed that the basis for survival is a sense of autonomy that people achieve by understanding the reasons for their

predicaments. How could that autonomy be promoted in a dictatorial and frightening regime?

Sutton's answer to this difficult question seems sensible. She does not deny or excuse the alleged behaviour, but she argues that Bettelheim nevertheless always showed a powerful ability to understand the problems of the children in his care and to promote that understanding in themselves. She gives us a picture, which I find convincing, of a powerful but fallible therapist.

This book is at its best in describing Bettelheim's reactions to his appalling experiences in Dachau and Buchenwald and to his work at the Orthogenic School. It is less impressive when the author comes to the ideas he developed on other topics. For example, Bettelheim's book on kibbutz children, *Children of the Dream*, provoked a great deal of controversy which Sutton describes very well. But she gives short shrift to the extremely interesting and sensitive account of the behaviour and the emotions of these children that makes up the core of the book. She mentions another book, *The Uses of Enchantment*, several times but she ignores its central argument: that children can use fairy tales as a metaphor to give them hope that there is a solution to their emotional problems. Again, the argument in another of Bettel-

heim's influential and pro-vocative books, *On Learning to Read* (written with Karen Zelan), is simply not mentioned by Sutton.

Nonetheless, this interesting and troubling biography of a brave and intelligent but deeply flawed psychologist is a considerable achievement. I hope it will encourage people to find out more about Bettelheim's ideas. □

Peter Bryant is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

Astrochemistry revealed

Stephen Lepp

The Chemically Controlled Cosmos: Astronomical Molecules from the Big Bang to Exploding Stars. By T. W. Hartquist and D. A. Williams. Cambridge University Press: 1995. Pp. 169. £19.95, \$39.95.

ASTRONOMERS study regions of the Universe that cannot be visited, inferring what they know about them from the light and particles reaching Earth. The most precise information comes in the form of light emitted by molecules at particular frequencies with an intensity dependent on the local environment, such as the density, temperature or radiation field. This makes such molecules excellent indicators of the physical conditions in astrophysical environments.

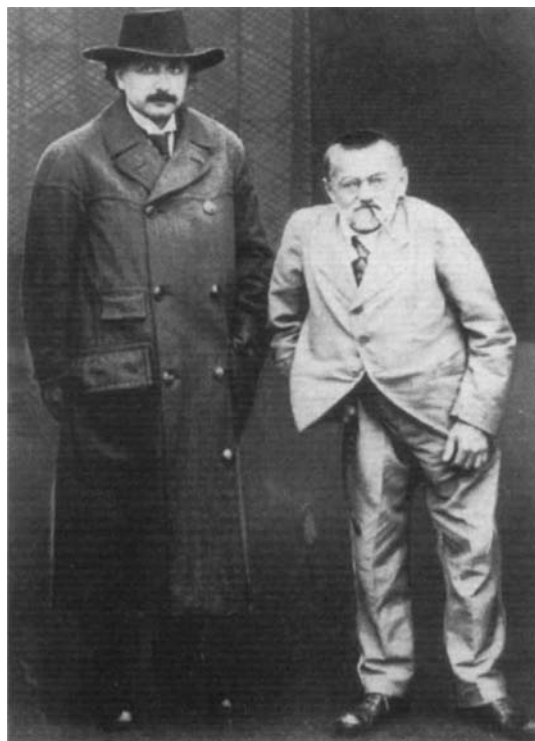
How molecules form and emit radiation is a subject that has long been neglected in the popular literature on astronomy. Hartquist and Williams have rectified the situation. As the title of their new book suggests, they not only explain how molecules may be used to probe the physical conditions, but they also show how molecules control the environment and evolution of astrophysical objects.

The presentation may at times be a bit difficult for a truly non-technical reader, although it should be accessible to anyone who regularly reads *Scientific American* or the News and Views section of *Nature*. It would make an excellent introduction to molecular astrophysics for a beginning graduate student, particularly if supplemented by the selected references at the end of each chapter. Astronomers, chemists or any scientist curious about the chemistry of astrophysical objects will find it a useful resource. I suspect most readers will be surprised at just how pervasive molecules are in the Universe, so much so that the book is nearly an overview of astronomy, covering such wide-ranging topics as the

formation of the first galaxies, star formation, stellar atmospheres, molecular clouds, supernovae, comets and active galaxies.

The largest part of the book is devoted to molecular clouds and how new stars are formed within them. Chemistry controls nearly all aspects of the birth of a star. Radiation from molecules cools the gas and sets the temperature, which in turn determines the mass of fragments that may collapse to form stars. The radiation also carries energy from the cloud, allowing the cloud to collapse. The chemistry determines the ionization fraction that couples the cloud to the magnetic field and impedes collapse. Molecules emitting light at radio and infrared frequencies also allow us to look inside regions of active star formation and help to solve the riddle of how stars are formed. Astrochemistry, controlling and revealing the cosmos, is now available to a wide audience. □

Stephen Lepp is in the Department of Physics, University of Nevada Las Vegas, 4505 Maryland Parkway, Box 454002, Las Vegas, Nevada 89154-4002, USA.



ALBERT Einstein with the American electrical engineer C. P. Steinmetz, c.1920. The picture is taken from *Essential Einstein* edited by A. B. Eddington. This slender glossy paperback contains 60 duotone photographs of Einstein, some famous, others less well known. Accompanied by quotations from his letters, speeches and recorded conversations, they provide a revealing portrait of the scientific, political and humane sides of the great man. Pomegranate, £22.95, £15.95.

Wide illumination

Frank McCapra

Photobiology. By E. Kohen, R. Santus and J. G. Hirschberg. Academic: 1995. Pp. 506. £54, \$69.95.

PHOTOBIOLOGY in its entirety is rarely given its proper place in university courses. The range of topics is great, and their study often leads to mutually exclusive specialization, despite the self-evident unifying presence of the action of light. As the authors of this book point out in their preface, photobiology is a truly multidisciplinary subject that touches on every aspect of life. A wide-ranging overview is therefore particularly valuable.

All aspects of the action of light on biological systems are covered here, but the treatment is uneven. Although the early chapters on the nature of light and some basic photophysics make for a convenient single-volume reference, similar and more detailed treatments of these topics can be found elsewhere. In any case, later cross-referencing to this background material is rarely provided, so these early chapters are a missed opportunity. What is more, details on the natural successor to such information — the photochemistry of relevant organic molecules — are almost wholly lacking. One might have expected, for example, a discussion in the chapter on vision of current thinking on *cis-trans* isomerism and models for the molecular basis of colour vision. (It seems that the authors were each responsible for isolated parts of the book and were self-effacing in referring to their own research.)

The chapters on vision and bioluminescence are disappointing: they fail to include the most up-to-date work and give insufficient details of the underlying organic chemistry. There is an effective emphasis on medical aspects, reflected by the large bibliographies for these chapters (in this respect some others in the book are seriously deficient), with many unpleasant, and not entirely necessary, photographs of diseases caused by or treated by light.

Nevertheless, the authors clearly bring out the importance of the interaction of light with the biosphere. They give a good picture of the enormous range of activity in this field and a fairly up-to-date description of discoveries and prospects in such research topics as environmental photobiology, photosynthesis, photocarcinogenesis and phototherapy. In short, it amounts to a reasonably successful, compact compendium. □

Frank McCapra is at the School of Chemistry, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

From DNA sequence to biological function

Stephen G. Oliver

Genome sequencing is leading to the discovery of new genes at a rate 50–100 times greater than that achieved by classical genetics, but the biological function of almost half of these genes is completely unknown. In order fully to exploit genome sequence data, a systematic approach to the discovery of gene function is required. Possible strategies are discussed here in the context of functional analysis in the yeast *Saccharomyces cerevisiae*, a model eukaryote whose genome sequence will soon be completed.

THE genome sequence of the human pathogen *Haemophilus influenzae* has recently been completed¹. Significant progress has also been made in sequencing the genomes of several model organisms, including those of the bacteria *Escherichia coli*² and *Bacillus subtilis*³, the budding yeast *Saccharomyces cerevisiae*⁴, the nematode *Caenorhabditis elegans*⁵, and the higher plant *Arabidopsis thaliana*⁶. In addition, work has begun on the genome of an important crop rice (*Oryza sativa*)⁷ and on the human genome itself⁸. The model organisms have been chosen, at least in part, because they have relatively small genomes (Table 1), allowing the sequencing of large chromosomal segments⁹ or, in the case of *S. cerevisiae*, entire chromosomes^{10–15}. This advantage is shared by rice. By contrast, the human genome of 3,000 megabases (Mb) is being studied mainly by sequencing complementary DNA copies of messenger RNA molecules (so-called ESTs, or expressed sequence tags¹⁶), although plans to make a high-resolution 'sequence map' have recently been announced¹⁷.

Sequencing chromosomal DNA can produce data unobtainable from the sequence of small genomic fragments or cDNA clones, including information on the relationship between the higher-order structure of chromosomes and such important events as recombination, transposition, replication and gene expression. What has caught the imagination of the general scientific community, however, is the speed with which large numbers of new genes are being revealed by systematic genome sequencing. The European Union Yeast Genome Sequencing Network alone is uncovering three new open reading-frames (ORFs) per day, about 40 times the rate achieved in the past 40 years¹⁸. It is already clear not only that the gene pool is much richer than we had anticipated, but also that the functions of nearly half the new genes are unknown. Moreover, this is as true of intensively studied organisms like *E. coli*¹ and *S. cerevisiae*³ as it is of more recent model organisms such as *Arabidopsis*⁵. Biologists are therefore faced with a major challenge in finding out what all these genes do.

Here I consider how this might be achieved for a simple model eukaryote, *S. cerevisiae*, and how such an exercise could assist the analysis of larger genomes.

Yeast as a lead organism for analysis

S. cerevisiae has a small genome (~13 Mb), a large number of chromosomes (sixteen), little repetitive DNA and few introns^{5,19}, features that have made it an excellent subject for genome sequencing. But it also has many features that commend it as a system with which to pioneer a systematic approach to screening for the function of novel genes. It is unicellular and can grow on chemically defined media, allowing the investigator complete control of its physiology. It can grow in either the haploid or the diploid state, so the effect of cell type or ploidy on gene action may be determined. Finally, and most importantly, excellent tools exist for its genetic manipulation, permitting the precise deletion of genes under investigation²⁰.

The problem of redundancy

Most of the *S. cerevisiae* genome is transcribed into RNA²¹, suggesting that it contains relatively little redundant information. This view was disturbed by an experiment²² in which the genome was disrupted at random, suggesting that about 70% of it was dispensable for growth on a rich, glucose-containing medium. The completion of the chromosome III sequence allowed a more specific analysis¹⁰: of 55 ORFs disrupted or deleted, only three were essential, and further analysis of 42 genes revealed a phenotype for only 21. Why does this small genome appear to have such a high level of redundancy?

Duplication of some important genes may provide a selective advantage by insuring against the loss of essential functions by mutation. Such genes may include those required for survival during stress or starvation. The genes encoding Hsp90 heat-shock proteins²³ and *RAS*²⁴ may be representatives of this class. Exact functional duplicates will be analysed most readily when the sequence of the entire yeast genome is complete and 'synthetic phenotypes', which result from multiple gene knockouts designed on the basis of sequence homology, can be examined.

Many yeast genes may be required to deal with challenges that are never encountered in the laboratory, but which are common in the rotting fig²⁵ or the brewery²⁶. In the laboratory, any of several closely related genes might be able to substitute for one or all of the others, but there may nevertheless be specific physiological challenges in which only one of the set can supply the exact function required for continued growth or survival. Yeast proteases furnish one example²⁷. A more specific example comes from chromosomes III and XIV, which apparently evolved from a common ancestor²⁸. Each has a citrate synthase gene at an equivalent position, but *CIT2* on chromosome III encodes the cytoplasmic enzyme, whereas *CIT1* on chromosome XIV encodes the mitochondrial enzyme (an apparent example of evolution through gene duplication).

It may also be that once the constraints on genome size imposed by the single, circular bacterial chromosome are released, organisms can afford to use multiple genes to perform identical (or nearly identical) functions in different contexts. Even in complex bacteria such as *Streptomyces* (which has a large, linear

TABLE 1 Sizes of genomes in current systematic sequencing projects

Organism	Genome size (Mb)
<i>Bacillus subtilis</i>	4.2
<i>Escherichia coli</i>	4.7
<i>Saccharomyces cerevisiae</i>	13.5
<i>Caenorhabditis elegans</i>	100
<i>Arabidopsis thaliana</i>	100
<i>Oryza sativa</i>	400
<i>Homo sapiens</i>	3,000

chromosome²⁹), multiple genes encoding enzymes of essentially identical biochemical function are used in secondary metabolism²⁸ and differentiation³⁰.

What physiological challenges need be analysed to reveal the functions of novel genes uncovered by sequencing? Obvious stresses such as heat, cold, osmotic shocks or starvation can be examined, but vast numbers of such conditions may be relevant to the natural history of yeast, many of them unknown. Another example from chromosome III illustrates the point. Although YCR32w (6,501 base pairs) is the longest ORF on the chromosome, its complete deletion had no obvious effect, and the mutant was exposed to many physiological challenges without results before it was discovered that it died when grown at low pH on glucose and challenged with acetic acid³¹. As YCR32w encodes a membrane protein, its product is probably an acetic acid exit pump. But can we reproduce this exercise for 3,000 more genes of unknown function once the yeast genome sequence is complete? A more systematic approach must be found.

If much redundancy is more apparent than real, the apparent rarity of gene deletions with a discernible phenotype may reflect a lack of adequate functional tests. There may thus be whole classes of functions to be found. But how can we uncover them? It is hard to believe that many undefined tracts of metabolism remain (although secondary metabolism in organisms other than antibiotic-producing bacteria is admittedly much neglected). New genes are therefore more likely to be involved in as-yet unstudied levels of regulation that may not be constitutively employed, such as the integration of different areas of metabolism, the temporal control of activity, or apoptosis. In order to reveal such functions, new methods of analysis will be needed; I now consider the strengths and weaknesses of several systematic approaches.

Analysis

Although the sequence of an ORF is itself a major resource that enables function to be inferred from the similarity that its putative protein product shows to other protein sequences in the data libraries, it is important to realize the limitations of such analyses. Significant matches of a novel ORF to another sequence may be in any of four classes. (1) A match that predicts both the biochemical and physiological function of the novel gene (such as that between ORF YCR24c from yeast chromosome III and *E. coli* Asn-tRNA synthetase). (2) A match that defines the biochemical function of a gene product without revealing its physiological function (such as the five new protein kinase genes found on chromosome III, whose biochemical function is clear—they phosphorylate proteins—but whose particular physiological function in yeast remains obscure). (3) A match to a gene from another organism whose function is unknown in that organism (such as that between ORF YCR63w from yeast chromosome III, protein G10 from *Xenopus*, and novel genes from *C. elegans* and humans, none of whose function is known). Such matches are becoming increasingly common as data accumulates from the different genomic sequencing projects. Although they illustrate the potential of systematic sequencing projects, they demonstrate that this potential may only be realized by an equally systematic experimental approach to functional analysis. (4) A match to a gene of known function that merely reveals that our understanding of that function is very superficial, such as that between YCL17c on yeast chromosome III and the NifS protein of nitrogen-fixing bacteria. In this instance, although similar genes were found in *Lactobacillus*, *Bacillus* and *E. coli*, neither yeast nor any of these other bacteria fix molecular nitrogen. Further studies in *Azotobacter*³² were necessary to show that NifS is responsible for inserting sulphur into the metal-sulphur centre of metalloenzymes, using pyridoxal phosphate as a cofactor. Whereas it may play an analogous role in *Bacillus subtilis*³¹, both sequence alignments^{33,34} and some experimental data³⁵ indicate that NifS proteins are pyridoxal-phosphate-dependent amino-transferases. Thus the discovery of similar genes in unexpected places has stimulated experiments that uncovered the true role of

a protein whose function was previously thought to be known. Similar matches may exist that do not seem incongruous to us and so do not immediately reveal the superficial nature of our knowledge.

Of course it is nice (and cheap) to think that the computer is the simplest route to function, but experiments remain essential. However, we cannot afford the time or money to perform all tests on all unknown genes, so a hierarchical taxonomy of gene function is needed that will allow us to perform only the tests relevant for genes of a given class (as in the 'keys' that are used as an aid to identifying species in classical taxonomy).

Yeast molecular genetics has recently been very successful in widening the net of gene interactions, identifying many new genes in the process. Two methods predominate: the isolation of successive extragenic suppressor mutations, and the creation of 'synthetic' phenotypes by combining two or more non-allelic mutations. The former can be applied to the current problem by cloning novel genes into multicopy vectors and examining their ability to suppress null mutations in known genes. The examination of all possible pairwise combinations is not really feasible, however, and a set of known genes whose suppression is likely to be particularly informative must therefore be chosen. The synthetic phenotype approach may also be helpful, but here too a battery of phenotypic tests will be required, as well as some hierarchical structure to guide their selective application, if it is not to become unwieldy. Similarity searches may make it obvious which combinations of deletions to analyse, but criteria other than sequence similarity may also be needed.

One way of classifying newly discovered proteins is to determine in which cellular compartment they are found. Burns *et al.*³⁶ have used mini-Tn3::lac Z to generate lac Z fusions with yeast ORFs in *E. coli*, producing fusion proteins that can be located by immunocytochemistry on transfer back into yeast. Of 2,373 fusions examined, 32% gave no staining, 58% gave general cytoplasmic staining, and 10% gave localized staining (Table 2). A parallel screen identified 39 fusions whose expression was specifically induced by meiosis.

Bolotin-Fukuhara is pursuing a complementary approach, in which a mini-Mu transposon is again used to make lac Z fusions³⁷. In this case, however, the lac Z fusion library is expressed in a pair of yeast hosts that differ only in the presence or absence of a particular transcriptional activator, allowing genes under the activator's control to be identified by their altered colour in the two strains. Although a similar approach was previously used to isolate co-regulated *E. coli* genes³⁸, this is probably the first application of the method to a eukaryote. A pilot experiment³⁷ identified 26 fusions regulated by the CCAAT-box-binding protein, Hap2p (refs 39, 40). This approach will be particularly helpful for new transcriptional activators identified by sequence analysis.

Classification

Virtually all the ORFs predicted by the chromosome III sequence are expressed by haploid cells growing on a rich glucose-based medium, at least as mRNAs^{41,42}. Nevertheless, expression levels are likely to vary greatly under different conditions, and a rapid procedure for northern analysis is required. This might entail hybridizing antisense oligonucleotide probes to a set of cDNA filters derived from cells (*MATa*, *MATx* and *MATa/MATx* diploid) grown under different conditions. A second, more direct approach is to establish a set of yeast ESTs⁴¹. Initial experiments have shown that the ESTs obtained from yeast grown on a minimal medium include many more genes of unknown function than those obtained from cells grown on rich medium⁴³. Careful cost-benefit analysis will be required to choose between the two approaches. Whichever is selected, it will be important to choose a small set of informative physiological conditions so as to simplify the grouping of genes according to function.

A related approach is based on two-dimensional protein gels,

TABLE 2 Screen for cellular location of protein products of yeast genes

Site of staining	No. of genes	No. of previously known genes	No. of sequenced genes not previously known	No. of unsequenced genes not previously known
Nucleus	98	11	3	7
Nucleolus	5			1
Endoplasmic reticulum/ nuclear rim	2		1	1
Punctate	143	3	34	
Cell periphery	10		1	1
Bud tip	2			1
Prospore wall	1			1

Previously known genes refer to database matches to the few genes for which partial sequence data is available where that matched was to a named yeast gene defined by 'classical' (function-first) genetic analysis. Not previously known genes refer to genes which have not been discovered by the classical route and whose sequence either is, or is not, currently available in the Genbank database. (This interpretation of the current situation was compiled from data supplied by Petra Ross-MacDonald.)

rather than northern blots. Although it is technically more complex, time-consuming and expensive than the northern approach, it can exploit deletion and amplification mutants, and can be combined with cell fractionation to demonstrate subcellular location (although cell fractionation with yeast is neither as efficient nor as discriminatory as the immunocytochemical approach described here⁴⁴). Most importantly, however, such analysis can reveal the set of proteins dependent on the expression of a given gene or that phosphorylated by a particular kinase.

Systematic phenotype screening

The phenotypic effect of deleting individual novel genes from yeast chromosome III has been examined in a qualitative but systematic manner. Tests were carried out on agar plates containing a large number of different growth media, sometimes incorporating specific metabolic inhibitors (K. Reiger *et al.*, manuscript submitted). Efficient methods of generating the required deletion mutants are now being developed^{45,46}, and it is hoped eventually to include deletions in all novel genes. Such a collection will be a resource of enormous importance.

Top-down metabolic control analysis

An approach that may be ideally suited to uncovering functional hierarchies rests on the principles of metabolic control (or 'flux') analysis^{47,48}. The two versions of this approach can be described as 'bottom-up' and 'top-down'. The former seeks to determine the contribution made by each enzyme in a metabolic pathway to the flux of carbon through that pathway. The analysis may be extended to enzymes more and more remote from the pathway being examined until all relevant enzymes have been identified and their contribution quantified. However, this approach is not applicable to the current problem, as the pathway affected by the gene under study is, by definition, unknown. The 'top-down' approach⁴⁹, by contrast, is admirably suited to the problem. Here, it is the flux through large metabolic domains that is measured. Smaller and smaller domains can then be examined, until the control of specific pathways or steps is elucidated. Top-down control analysis, therefore, has exactly the type of hierarchical structure necessary to facilitate the search for gene function, and the concepts involved can readily be applied to spheres of biological activity other than metabolism.

Conclusions and prospects

Systematic genome sequencing presents biology with enormous opportunities, and some dangers. The dangers include the mistaken belief that analysis *in silico* will by itself be sufficient to reveal the function of all novel genes uncovered. An equally wrong-headed alternative is to regard the genome sequences as vast reference works whose purpose is simply to save researchers the bother of sequencing the genes on which they alight in the course of their investigations. Seductive though this view is, there may be good reasons why classical genetics has not discovered these genes, and we should not be so arrogant as to imagine that

there are no new functions left to discover. Many genes of unknown function are under-represented in cDNA libraries (R. Waterston, personal communication), indicating that they are only expressed transiently or at very low levels, perhaps because they play a regulatory role. Even if they should eventually be revealed by current approaches, this is an enormous waste of the hard-won resource the sequences represent, which could instead be used to transform the pace of biological research.

Although we do not understand the function of about half the genes in most organisms, we have a good understanding of the function of the other half in organisms such as yeast and *E. coli*. Like the sequences themselves, this is an enormous resource which should be exploited in any systematic scheme for functional analysis. The known genes may be used to optimize the tests employed in our function searches, and double-blind trials can be carried out to ask whether new schemes can correctly classify the functions of known genes. Moreover, metabolic control analysis suggests that the overexpression of particular genes of known function will permit the construction of a set of strains that will be sensitized to the detection of the effects of novel genes in specific domains of biological activity.

S. cerevisiae has a special role to play in this process. The function of yeast's 7,000 or so genes, once known, will provide a working description of a eukaryotic cell. It will also provide a unique tool with which to search for drugs to treat not only infectious diseases but also major killers such as cancer⁵⁰.

Analysis of the yeast genome is likely to contribute in several ways to the human genome project⁵¹. It will assist in identifying the function of newly discovered human genes, both by sequence comparisons, and by complementation of defined genetic lesions in yeast with human cDNA clones. This process might be facilitated by the construction of a 'minimalist yeast', a strain in which all apparently redundant gene sets are reduced to the minimal set permitting growth on complex media in the laboratory. The remaining genes may then be successively replaced by human or viral sequences. As the viability of such a strain would then be absolutely dependent on the heterologous gene, it would form a powerful tool for screening antitumour and antiviral drugs.

The analysis of the yeast genome will also assist studies of proteins implicated in human heritable diseases. Many such proteins have yeast homologues^{52,53} (Table 3) and the study of their physiological role and their interactions with other yeast proteins will facilitate understanding of the corresponding disease. Many important medical conditions in humans, whether actual diseases (such as early-onset diabetes⁵⁴) or a predisposition to them (such as colon cancer⁵⁵ or heart disease^{56,57}) are controlled by multiple genes⁵⁸, and uncovering the complete set of genes involved in a particular condition is a difficult and lengthy process⁵⁴⁻⁵⁸. Recognition of homology between a yeast protein and a human protein implicated in a particular disease may thus provide an important key to improvements in diagnosis or therapy. *In vivo* approaches to identify interactions between

TABLE 3 Sequence similarity between positionally cloned human genes and *S. cerevisiae* ORFs discovered by systematic sequencing

Human disease	MIM no.	Human gene	GenBank no. for human cDNA	BLASTX P-value	Yeast gene	GenBank no. for yeast DNA
Lowe's syndrome	309000	OCRL	M88162	1.2e-47	YIL002c	Z47047
Breast cancer, early onset	600185	BRCA2	U43746	4.3e-4	YER033c	U18796
Neurofibromatosis, type 2	101000	NF2	L11353	4.3e-4	N2231	X85811
Kallmann's syndrome	308700	KAL	M97252	1.0e-3	YKL083w	Z28082
Tuberous sclerosis	191090	TSC	X75621	1.5e-3	ORF00954	X83121
Marfan's syndrome	154700	FBN1	L13923	1.9e-2	YCR89w	X59720
Huntington's disease	143100	HD	L12392	3.4e-2	D4411	X82086
Long Q-T syndrome, type 1	192500	KVLQT1	U40990	3.1e-1	YBR235w	Z36104
Fragile-X syndrome	309550	FMR1	S65791	4.1e-1	UND407	U43491
Emery-Dreifuss muscular dystrophy	310300	STA	X82434	5.7e-1	YBL046W	Z35807
Norrie's disease	310600	NDP	X65882	9.4e-1	YIL037c	Z47047

BLASTX searches were performed using each positionally cloned human cDNA listed against a non-redundant database of *S. cerevisiae* sequences (NRSC) maintained at the Saccharomyces Genomic Information Resource. For the most statistically significant yeast protein match for each human query, the BLAST (v1.4) *P*-value for the match is given together with the GenBank accession number for the corresponding *S. cerevisiae* genomic sequence. Data shown represent a subset of a more complete analysis (D. E. Bassett Jr which is accessible on the World-Wide Web (<http://www.ncbi.nlm.nih.gov/Bassett/Yeast/PosiCloneSceNew.html>)). Further details of this analysis and its results are given in Scientific Correspondence by D. E. Bassett *et al.* in this issue.

genes or their products (such as the identification of extragenic suppressors, control-of-expression studies³⁸, and the use of the two-hybrid system⁵⁰) may also reveal other proteins whose human homologues are involved in the disease.

It is clear that the functions of many novel genes discovered by DNA sequencing will be uncovered and set in their correct physiological context over the next few years. This process will

be most important for human genes, but the search for function in the genes of *S. cerevisiae* will undoubtedly be an essential navigation aid for what is to follow. □

Stephen Oliver is in the Department of Biochemistry and Applied Molecular Biology, UMIST, PO Box 88, Manchester M60 1QD, UK.

- Fleischmann, R. D. *et al.* *Science* **269**, 538–540 (1995).
- Blattner, F., Daniels, D. L., Burland, V. D., Plunkett, G. & Chang, S. in *The Chromosome* (eds Heslop-Harrison, J. S. & Flavell, R. B.) 43–59 (Bios, Oxford, 1993).
- Devine, K. *Trends Biotechnol.* **13**, 210–216 (1995).
- Oliver, S. G., James, C. M., Gent, M. E. & Indge, K. J. in *The Chromosome* (eds Heslop-Harrison, J. S. & Flavell, R. B.) 233–248 (Bios, Oxford, 1993).
- Sulston, J. *et al.* *Nature* **356**, 37–41 (1992).
- Schmidt, R. & Dean, C. *Bioessays* **15**, 63–69 (1993).
- Havukkala, I., Ichimura, H., Nagamura, Y. & Sasaki, T. *J. Biotech.* **41**, 139–148 (1995).
- Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4338–4344 (1993).
- Wilson, R. *et al.* *Nature* **368**, 32–38 (1994).
- Oliver, S. G. *et al.* *Nature* **357**, 38–46 (1992).
- Dujon, B. *et al.* *Nature* **369**, 371–378 (1994).
- Johnston, M. *et al.* *Science* **265**, 2077–2082 (1994).
- Feldmann, H. *et al.* *EMBO J.* **13**, 5795–5809 (1994).
- Bussey, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **92**, 3809–3813 (1995).
- Murkamai, Y. *et al.* *Nature Genet.* **10**, 261–268 (1995).
- Adams, M. D. *et al.* *Science* **252**, 1652–1656 (1991).
- Waterston, R. & Sulston, J. *Nature* **376**, 111 (1995).
- Olson, M. V. in *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae* (eds Broach, J. R., Pringle, J. R. & Jones, E. W.) 1–39 (Cold Spring Harbor Laboratory, New York, 1991).
- Rothstein, R. *J. Meth. Enzym.* **101**, 202–211 (1983).
- Kaback, D. B., Angerer, L. M. & Davidson, N. *Nucleic Acids Res.* **6**, 2499–2517 (1979).
- Goebel, M. E. & Petes, T. D. *Cell* **46**, 983–992 (1986).
- Craig, E. A., Gambill, B. D. & Nelson, R. J. *Microbiol. Rev.* **57**, 402–413 (1993).
- Tatchell, K., Robinson, L. C. & Breitenbach, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3785–3789 (1985).
- Mortimer, R. K. & Johnston, J. R. *Genetics* **113**, 35–43 (1989).
- Hammond, J. R. M. in *The Yeasts* Vol. 5 (eds Rose, A. H. & Harrison, J. S.) 7–67 (Academic, London, 1993).
- Renduelles, P. S. & Wolf, D. H. *FEMS Microbiol. Rev.* **54**, 17–46 (1988).
- Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. & Thuriaux, P. *Compt. Rend. Acad. Sci. (III)* **316**, 137–143 (1993).
- Lin, Y. S., Kieser, H. M., Hopwood, D. A. & Chen, C. W. *Molec. Microbiol.* **10**, 923–933 (1993).
- Chadwick, D. J. & Whelan, J. (eds) *Secondary Metabolites: Their Function and Evolution* (CIBA Foundation Symposium 171), (Wiley, Chichester, 1992).
- Chater, K. F. in *Regulation of Prokaryotic Development* (eds Smith, I., Stepecky, R. A. & Setlow, P.) 277–299 (Am. Soc. Microbiol., Washington, 1989).
- Jia, Y. thesis, Univ. Pierre et Marie Curie, Paris (1993).
- Zheng, L., White, R. H., Cash, V. L., Jack, R. F. & Dean, D. R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2754–2758 (1993).
- Sun, D. & Setlow, P. *J. Bact.* **175**, 1423–1432 (1993).
- Mehta, P. K. & Christen, P. *Eur. J. Biochem.* **211**, 373–376 (1993).
- Ouzunis, C. & Sander, C. *FEBS Lett.* **322**, 159–164 (1993).
- Leong-Morgenthaler, P., Oliver, S. G., Hottinger, H. & Sölli, D. *Biochimie* **76**, 45–49 (1994).
- Burns, N. *et al.* *Genes Dev.* **8**, 1087–1105 (1994).
- Dang, V.-D., Valens, M., Bolotin-Fukuhara, M. & Daignan-Fornier, B. *Yeast* **10**, 1273–1283 (1994).
- Casadaban, M. J. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4530–4533 (1979).
- Olesen, J., Hahn, S. & Guarente, L. *Cell* **51**, 953–961 (1987).
- De Winde, J. H. & Grivell, L. A. *Progr. nucleic Acids Res.* **46**, 51–91 (1992).
- Yoshikawa, A. & Isono, K. *Yeast* **6**, 383–401 (1990).
- Tanaka, S. & Isono, K. *Nucleic Acids Res.* **21**, 1149–1153 (1992).
- Weinstock, K. G., Kirkness, E. F., Lee, N. H., Earle-Hughes, J. A. & Venter, J. C. *Curr. Opin. Biotech.* **5**, 599–603 (1994).
- Kreutzfeldt, C. & Witt, W. in *Saccharomyces* (eds Tuite, M. F. & Oliver, S. G.) *Biotech. Handbooks* Vol. **4**, 5–58 (Plenum, New York, 1991).
- Rieger, K., Orłowska, G., Kaniak, A., Aljinovic, G. & Slonimski, P. *Yeast* (in the press).
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21**, 3329–3330 (1993).
- Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P. *Yeast* **10**, 1793–1808 (1995).
- Kacser, H. & Burns, J. A. *Symp. Soc. exp. Biol.* **32**, 65–104 (1973).
- Heinrich, R. & Rapoport, T. A. *Eur. J. Biochem.* **42**, 89–95 (1974).
- Quant, P. A. *Trends biochem. Sci.* **18**, 26–30 (1993).
- Dulbecco, R. *Gene* **135**, 259–260 (1993).
- Short, N. *Nature* **377** (suppl.), 1 (1995).
- Tugendreich, S., Bassett, D. E. Jr, McKusick, V. A., Boguski, M. S. & Hieter, P. *Human molec. Genet.* **3**, 1509–1517 (1994).
- Bassett, D. E. Jr, Boguski, M. S. & Hieter, P. *Nature* **379**, 589–590 (1996).
- Davies, J. L. *et al.* *Nature* **371**, 130–136 (1994).
- Bodmer, W., Bishop, T. & Karan, P. *Nature Genet.* **6**, 217–219 (1994).
- Ward, R. in *Hypertension: Pathophysiology, Diagnosis and Management* (eds Laragh, J. H. & Brenner, B. M.) 81–100 (Raven, New York, 1990).
- Schwartz, K. *Nature Genet.* **8**, 110–111 (1994).
- Lander, E. S. & Schork, N. J. *Science* **265**, 2037–2048 (1994).
- Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9578–9582 (1991).

ACKNOWLEDGEMENTS. Work on yeast genome sequencing and analysis in my own laboratory is supported by the European Commission, BBSRC, and by Pfizer Central Research. I would like to thank J. Fincham for asking the right questions; H. Boucherie, B. Dujon, S. Fey, P. Hieter, P. Mose Larsen and M. Snyder for critically reading the manuscript; K. Indge for his help with computer analyses; and D. Bassett, A. Goffeau, P. Slonimski and P. Ross-MacDonald for their free communication of data.

Dynamical weakening of faults by acoustic fluidization

H. J. Melosh

Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

When a fault slips seismically, some of the energy released may excite strong, short-wavelength vibrations near the fault core. Such vibrations can temporarily reduce the normal stress on the fault, allowing it to slip at lower shear stresses than predicted by laboratory coefficients of friction. This phenomenon of 'acoustic fluidization' provides an alternative to theories that invoke pressurized fluids as an explanation for why some faults appear to be so weak.

RESEARCH over the past two decades has confirmed the early suggestion of Brune *et al.*^{1,2} that the rate of frictional heat generation, and by inference, the overall driving stress on the San Andreas fault, is substantially less than would be expected from the known slip rate and laboratory measurements on rock friction. Modern heat flow measurements still fail to detect a significant anomaly connected with the fault^{3,4}, and new data on both *in situ* stresses^{5,6} and the orientation of geological structures in the vicinity of the fault^{7,8} indicate that the resolved shear stress on the fault is smaller than 10–20 MPa, nearly an order of magnitude smaller than the normal (perpendicular) stress expected on a 10-km-deep fault (150 MPa, at mid-depth). Despite many attempts to explain this apparent weakness otherwise, most geoscientists now seem convinced that the San Andreas, and by extension many other major faults, exhibits an astonishingly low resistance to shear stresses⁹, although other, perhaps less mature, faults do seem to support higher stresses¹⁰. Laboratory measurements on both intact and ruptured rock specimens have not yet offered any explanation for this very low apparent strength of fault zones, and the disagreement between the observed low strength in the field and the properties of rock specimens in the laboratory has attained the status of a geophysical paradox.

In the past few years several groups have offered explanations of this low strength by a variety of mechanisms that all appeal to the presence of water in the fault zone. Sleep and Blanpied¹¹ and Byerlee¹² both suggested models in which high-pressure water is trapped in the fault zone. Rice¹³ suggested that water from the lower crust may preferentially pressurize a high-permeability fault zone, leading to localized failure at low shear stress. Unfortunately, recent measurements of porosity changes in shearing rocks¹⁴ indicate that the increase of volume during shear failure overshadows the pressure increase due to any putative compaction, so that these proposed weakening mechanisms may not work.

In 1979 I proposed a weakening mechanism that relies on the development of strong, high-frequency vibrations near the fault zone¹⁵. This mechanism, which I called "acoustic fluidization" because it operates through the action of strong elastic (acoustic) waves in the surrounding rocks, depends on a temporary reduction of normal stresses by these waves, so that sliding may temporarily (and locally) take place under much lower shear stresses than required in the absence of the vibrations. Water or other fluids are not necessary for this process. There is currently a new emphasis on the role of normal vibrations (perpendicular to the fault plane) in reducing friction^{16–19}, and it seems worthwhile to elaborate my initial proposal by further developing the equation describing the balance of acoustic energy and relating that balance to the strain rate in the core of a fault. I apply this equation to a

planar fault and show that, not only do plausible assumptions about scattering and absorption of acoustic energy in the Earth yield an order of magnitude reduction in the coefficient of friction, but that the nonlinear equation possess two solutions, one corresponding to rapid seismic slip and the other to a much slower motion, perhaps offering an explanation for aseismic creep or 'silent' earthquakes.

Fundamental equation of acoustic fluidization

The basic idea of the acoustic fluidization mechanism is that, although the seismic rupture may appear to be a simple shear displacement from a distance (via focal mechanism studies), the propagation of the rupture in detail is irregular and noisy. Some fraction of the total energy is released as short-wavelength elastic waves that are scattered by small-scale heterogeneities in the fault zone, and hence constantly change mode and direction. Some of these waves generate oscillations in the normal stress across the fault zone, facilitating failure under average shear stresses that would not be large enough to cause sliding under static conditions.

For this mechanism to operate, the magnitude of the stress fluctuations must approach the overburden stress on the fault, ρgh where ρ is the density of the crustal rocks, g is the acceleration of gravity, and h is the depth of the fault (I ignore small effects due to depth-dependent density). The elastic energy, per unit volume, in a wave of this amplitude is of the order of $(\rho gh)^2 / 2\rho c^2$, where c is the velocity of sound in the elastic material.

Simple energy balance requires that these high-amplitude vibrations are localized in a relatively narrow region near the fault. Thus the total elastic energy of the vibrations must be much less than the total energy release τu in the earthquake itself, where τ is the resolved shear stress on the fault, u is the total slip and the energy is normalized by the fault area.

In my 1979 paper I estimated that the width of this fluidized zone must be less than about 1 m. However, in that paper I chose a value for c that is a factor of 2 smaller than the typical crustal shear velocity of about 4 km s⁻¹. Furthermore, I assumed that only 10% of the energy release was in the form of elastic waves, and this now seems overly pessimistic. Thus my original estimate that a 1-m-wide zone near the fault can be fluidized should now probably be regarded as a lower limit, the actual width being as much as 20 m. This requirement is easily satisfied within the cores of observed fault zones, where the central ultracataclasite zone may be only centimetres to a few tens of centimetres thick²⁰, and adjacent sheared regions are of the order of a metre thick.

As the waves must be confined in a zone of this width, or less, their wavelengths must accordingly be shorter than a few metres. These high-amplitude and high-frequency waves do not propagate far from the fault zone (see below) and are thus difficult to observe

unless an instrument with a high frequency response (1–10 kHz) is present very close to the rupture surface.

Satisfying energy conservation, however, does not guarantee the presence or effectiveness of short-wavelength vibrations. It remains to show that an appropriate feedback can exist between the generation of such vibrations and their facilitation of the larger-scale flow, which in turn regenerates more of the vibrations, all at shear stress levels well below that given by normal coefficients of rock friction²¹

The first step in establishing such a relationship is to determine how the amplitude of strong vibrations affects the rate of slip, or strain rate $\dot{\epsilon}$ of rock debris. Assuming that the internal gouge zone in a mature fault core lacks significant cohesion, the strain rate is given by the strain change in each failure event, divided by the period of the vibrations, times the probability (per period) that failure will occur. At a particular wavelength λ , this is $\dot{\epsilon} = (\text{strain change/period})(\text{probability of failure})$ or, using the quantities defined previously $\dot{\epsilon} = [\tau/(\rho\lambda c)]P(\text{stress} > \rho gh)$. For the special case of gaussian random noise this expression has the form¹⁵

$$\dot{\epsilon} = \frac{\tau}{\rho\lambda c} \left\{ \frac{1 - \text{erf}(\rho gh/\sqrt{2}\sigma)}{1 + \text{erf}(\rho gh/\sqrt{2}\sigma)} \right\} \quad (1)$$

where σ is the variance of the pressure fluctuation spectrum and erf is the error function. Other statistical distributions of pressure fluctuations are possible, or even likely, but this form is widely used in the engineering analysis of vibrations²² and it serves as a useful prototype until direct measurements of such vibrations along faults are possible. This equation is slightly different from the 1979 paper because I have dropped a term $-\tau/\mu$ in the argument of the error function. This is a good approximation if τ is much less than ρgh , as will be shown. I omit this term because it adds an inessential new degree of freedom to the equations. In any event the simplification is conservative: the actual strain rates are somewhat higher than equation (1) would suggest, especially as τ approaches the overburden stress.

Although many geophysicists were initially sceptical that vibrations could make granular materials flow like liquids, there is now ample evidence for the existence of this phenomenon from both engineering practice, such as vibratory pile driving²³, as well as from shake tables²⁴ and the response of factory foundations to strongly vibrating machinery²⁵. Some carefully controlled observations of this phenomenon now exist²⁶. Most of these engineering

observations take place at much lower frequencies than necessary for an explanation of seismic slip, and all are strongly affected by the presence of a free surface, but they illustrate the basic principle that vibrations can drastically change the rheology of loose granular materials. Vibrations thus turn materials that behave like coulomb solids (one whose strength increases with pressure) into non-newtonian viscous fluids²⁷.

The second, and crucial, step in making feedback possible is to note that when a region that is elastically strained fails owing to a sudden decrease in the local overburden stress, its internally stored elastic energy is released and a fraction, e , is available to regenerate the vibrations. This "seismic efficiency" e is typically estimated²⁸ at about 0.5. As the total rate of energy release per unit volume is $\dot{\epsilon}\tau/2$, the volumetric rate of production of elastic wave energy during this vibrationally enhanced flow is $e\dot{\epsilon}\tau/2$.

So far, nothing in this discussion has indicated how the wavelength of these vibrations is determined, other than by the energetic requirement that it be short (metres or less). It cannot be shorter than the grain size, otherwise the continuum description would fail and the elastic energy could not propagate as a wave. The process of energy regeneration probably does not directly establish a wavelength either: the overburden pressure is relieved nearly uniformly over a volume with a diameter of roughly half a wavelength. The strained material in this volume rebounds elastically, radiating elastic energy at nearly the same wavelength. In other words, energy regeneration is a near-linear process that is not very effective in converting one wavelength to another. However, the overall requirements of regeneration may determine a wavelength by a kind of natural selection. Thus, those wavelengths for which regeneration is particularly strong will get stronger, while wavelengths where absorption or radiation (which is particularly important at long wavelengths) dominates will become relatively weaker. Just how (and whether) this process works is unclear and may require a combination of *in situ* study and perhaps numerical modelling to fully resolve. For the present, parametrizing the feedback by an efficiency factor e is probably the best that can be done.

A complete analysis of the energy budget must also take into account the loss of elastic wave energy to heat and to propagation either into or away from the fault zone. To facilitate this analysis I assume that short-wavelength elastic wave propagation is dominated by scattering. In this case the most useful descriptor of the wave energy is the elastic energy density, whose propagation can be described by a diffusion equation (first used to analyse the propagation of seismic energy on the Moon^{29,31} where scattering dominates even normal seismograms owing to the extremely low dissipation). This approximation is best thought of as an end-member of a spectrum of elastic energy propagation modes. At the other end of this spectrum is Brune's normal-fluctuation model¹⁶ where normal stresses are excited by a wave coherent with the slip phase. The short wavelengths of the vibrations near the fault, however, make strong scattering likely in the incoherent, shattered rock observed close to the cores of mature faults²⁰. Furthermore, recent work has shown that a substantial amount of energy is scattered into the coda of even normal seismic waves. The energy decline in these codas is also well described by a diffusion-type equation³².

To formally analyse the energy balance of the short-wavelength vibrations, consider a control volume near the fault which is small compared to the fault itself, but large compared to the wavelength. Elastic wave energy can be scattered either into or out of the volume. Once inside, it can be absorbed as heat, or it can facilitate failure of loose material in the fault zone and generate new elastic energy. The full equation describing these possibilities is

$$\frac{dE}{dt} = \frac{\xi}{4} \nabla^2 E - \frac{c}{\lambda Q} E + e \frac{\dot{\epsilon}\tau}{2} \quad (2)$$

The left-hand side of the equation describes the rate of change of the elastic energy density E in the control volume. The derivative is, strictly speaking, a convective material derivative,

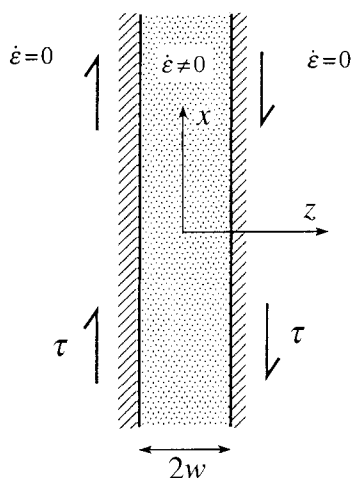


FIG. 1 Schematic illustration of the fault zone geometry in which a steady-state solution is sought. The scattering properties of the fault core and walls are identical, the only difference is that the strain rate $\dot{\epsilon}$ is set to zero in the walls, whereas the fault core is assumed to be broken, non-cohesive rock debris that flows if the yield stress is exceeded. w is the half-width of the fault and τ is the applied shear stress.

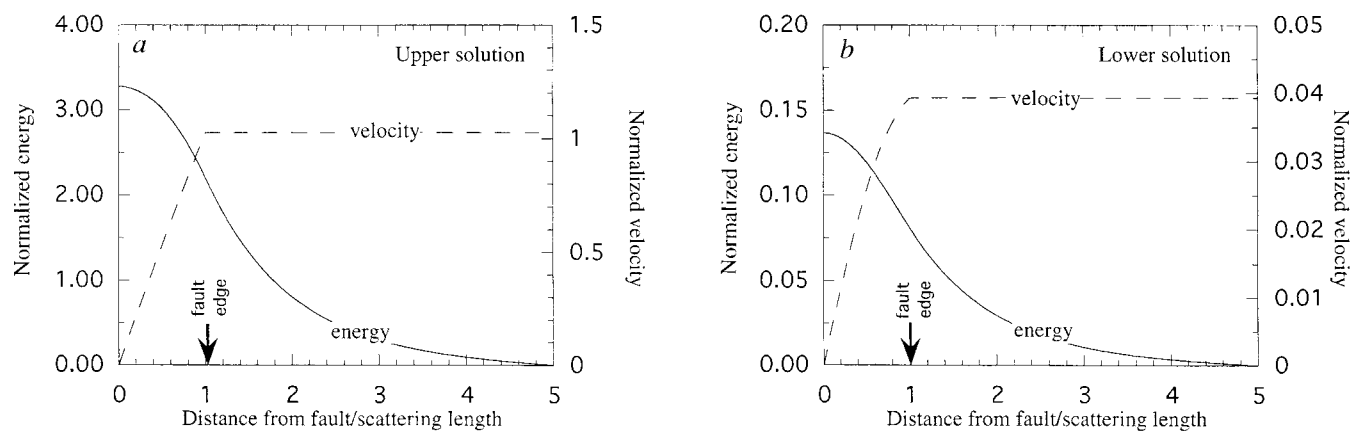


FIG. 2 *a*, Steady-state solution on the upper or 'fast' branch of solutions. Note that acoustic energy fills the fault zone and that the energy density falls exponentially in the non-shearing walls of the fault. The solid lines indicate energy density (left scale), the dashed lines indicate velocity (right scale). Both are normalized as discussed in the text. The regeneration

parameter $r\Sigma^2 = 10$ in this computation, and the fault zone half-width is 1. *b*, Steady-state solution on the lower or 'slow' branch of solutions. Note that the acoustic energy is somewhat more tightly confined than in *a*, and that both energy density and velocity are much lower. The regeneration parameter $r\Sigma^2 = 10$ in this computation.

but as material velocities during earthquakes are much less than the sound velocity this distinction can be ignored. The first term on the right describes scattering into or out of the control volume, where ξ is a scattering diffusivity, which has dimensions of a length times a velocity. The second term parametrizes conversion of elastic energy into heat by the inverse intrinsic dissipation, Q . These terms together are linear in E , and if they were the only ones present the solutions of equation (2) would be straightforward but uninteresting. The third term represents energy generation during flow subjected to an (assumed) constant tectonic shear stress τ . It is a highly nonlinear function of E because the variance σ of the vibrations in equation (1) depends on E as $\sigma = c\sqrt{2\rho E}$ for gaussian noise. This term leads to far more interesting behaviour.

Application to a planar fault

The fundamental equation describing the propagation, absorption and generation of elastic energy in the limit of strong scattering (equation (2)) is a nonlinear partial differential equation that depends on three spatial variables and time. It is thus extremely difficult to solve without simplifying approximations. Because the main goal of this Article is to demonstrate the existence of solutions to this equation that correspond to fault slip at stresses much lower than those derived from a normal coefficient of static friction, it is enough to seek steady-state solutions for slip on a planar fault at constant velocity. Time-dependent solutions describing the onset or termination of seismic slip will be treated elsewhere.

Thus, for a steady-state solution, the left-hand side of equation (2) is zero: $dE/dt = 0$. As the fault is assumed to be planar, all spatial derivatives in the direction of the fault plane also vanish. If z is the distance normal to the fault plane, then the only non-zero derivatives are with respect of z . In this case equation (2) simplifies to:

$$0 = \frac{\xi}{4} \frac{d^2 E}{dz^2} - \frac{c}{\lambda Q} E + e \frac{\dot{\epsilon} \tau}{2} \quad (3)$$

using equation (1) for gaussian random noise, the strain rate is given by

$$\dot{\epsilon} = \frac{\tau}{\rho \lambda c} \left\{ \frac{1 - \text{erf}(x)}{1 + \text{erf}(x)} \right\} \quad (4)$$

in which $x = \rho g h / 2c \sqrt{\rho E}$.

This equation can be further simplified by normalizing the

various terms by constant quantities of the same dimensions and some convenient size. Thus, the energy density E is normalized by the energy density at which the pressure fluctuations approximately equal the overburden stress at the fault depth,

$$\Psi = \frac{E}{\left\{ \frac{(\rho g h)^2}{\rho c^2} \right\}} \quad (5)$$

Distances are normalized by the "scattering length", l_s , which is the characteristic distance over which scattering and dissipation reduce the amplitude of the elastic energy by $1/e$: $\zeta = z/l_s$, where $l_s = \sqrt{\xi \lambda Q / 4c}$.

Stresses are normalized by the overburden stress at the bottom of the fault, given by $\Sigma = \tau / \rho g h$. In terms of these non-dimensional variables the combinations of equations (3) and (4) simplify to

$$\frac{d^2 \Psi}{d\zeta^2} = \Psi - r \Sigma^2 \left[\frac{1 - \text{erf}(1/2\sqrt{\Psi})}{1 + \text{erf}(1/2\sqrt{\Psi})} \right] \quad (6)$$

There is only a single constant in this equation, $r\Sigma^2$, which thus controls the nature of all solutions. This term is derived from the energy-feedback term in equation (2). If this term is zero, then there are no interesting solutions. This term depends on the square of the applied shear stress Σ , and on the factor $r = eQ/2$, which I call the "regeneration" parameter.

Figure 1 illustrates the geometry in which equation (6) is solved. The fault is idealized as a narrow core of broken, non-cohesive rock of half-width w . Outside this core the walls of adjacent rock are assumed to have the same scattering properties (this is for convenience only: it is easy to assume different properties, but there is no particular reason to do so, nor does it change the character of the solutions). However, the adjacent rock is assumed to be cohesive, so $\dot{\epsilon} = 0$ outside the fault core and no elastic energy regeneration occurs. In this region the equation for the energy density is simply $d^2 \Psi / d\zeta^2 = \Psi$ for $z > w$ where equation (6) applies for $z < w$. Solutions are obtained by matching both the energy density Ψ and its first derivative at $z = w$. The highly nonlinear system of equations is solved numerically using a fourth-order Runge-Kutta integrator and a bisectional shooting method²³ that starts with an assumed energy density at the centre of the fault, $\Psi(0)$ ($d\Psi/d\zeta(0) = 0$ by symmetry), then integrates outward beyond the fault core and away from the fault. A solution

is sought for which $\Psi \rightarrow 0$ as $\zeta \rightarrow \infty$. When a solution is found for the energy density, the non-dimensional velocity of the flow v can be determined from the strain rate by integration (recall that $\dot{\epsilon}_{zx} = 0.5(dv_x/dz)$ where x is in the direction of slip). Thus, the velocity is determined from equation (1)

$$\frac{dv}{d\zeta} = 2 \left[\frac{1 - \operatorname{erf}(1/2\sqrt{\Psi})}{1 + \operatorname{erf}(1/2\sqrt{\Psi})} \right] \quad (7)$$

where v is determined from the slip velocity $\dot{u}$ by

$$v = \frac{\dot{u}}{\frac{\tau}{\rho c} \frac{l_s}{\lambda}} \quad (8)$$

This equation holds only in the deforming zone, $z < w$. For $z > w$, v is constant because no strain is permitted to accumulate there.

An example of one of these solutions is shown in Fig. 2a. In this solution the fault core extends ± 1 scattering length from its centre, and the product $r\Sigma^2 = 10$. It is clear that a non-trivial solution exists in this case, in which the energy density in the core is close to unity (that is, the amplitude of the pressure fluctuations is comparable to the overburden stress), and a substantial velocity exists across the fault. The fault core is nearly filled with powerful elastic vibrations, and the fact that a steady-state solution has been found means that large strains can accumulate.

A non-trivial solution cannot always be found: if the parameter $r\Sigma^2$ is too small, then no solution other than $\Psi = 0$ exists to these equations. In other words, if regeneration is too weak, absorption too strong, or the driving stress Σ too small, then no acoustically fluidized slip is possible. The minimum value of $r\Sigma^2$ depends somewhat upon the fault zone width w , but for w close to 1 or larger, the minimum $r\Sigma^2$ is 2.8. This value represents a critical threshold. Such threshold behaviour is common in nonlinear differential equations, where such points are often associated with bifurcations³⁴. Equation (6) is no different: above the threshold value of 2.8 there are two solutions.

Figure 3 shows the characteristics of solutions to the fault problem as a function of normalized central energy density $\Psi(0)$ and the parameter $r\Sigma^2$ for a variety of fault widths. The upper solutions with $\Psi(0) > 1$ correspond to the solution illustrated in Fig. 2a. It can be shown that they are stable to small perturbations in elastic density for constant shear stress. Figure 3 also shows a branch of solutions with $\Psi(0) \ll 1$, Fig. 2b illustrates one of these solutions. These solutions have much smaller energy densities and are self-localizing, in that the zone of large energy density may be only a small fraction of the width of the fault core. Because energy densities are low, the slip velocities are correspondingly low. I will argue later that this class of steady, very low-energy, low-velocity solutions may correspond to either fault creep episodes or 'silent' earthquakes.

Application to the Earth

The equations for acoustic energy generation and decay have mathematical solutions that resemble motion observed in earthquakes. Steady slip is possible under conditions of constant applied shear stress because strong vibrations in the fault core permit the material there to flow like a viscous liquid. At the same time, a fraction of the energy dissipated in the flow makes up for losses in the elastic vibrations. However, do these solutions solve the principal paradox of slip on major faults? That is, do these solutions represent slip under low resolved shear stress? To answer this question real numbers must be substituted for some of the parameters in the theory. The most important parameter is $r\Sigma^2$, which was shown to have a minimum value of about 2.8 in the last section (see also Fig. 3). Because the stress Σ is normalized by the overburden pressure, a value of Σ much less than 1 corresponds to flow at low stress. Observations on the San Andreas fault suggest that Σ on that fault may be as low as 0.1. For solutions to exist for this value of Σ , the regeneration parameter $r = eQ/2$,

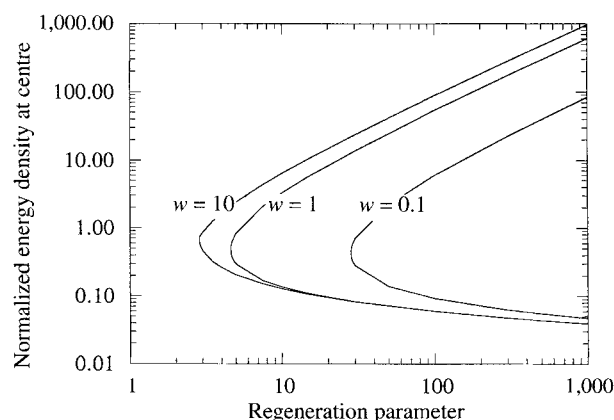


FIG. 3 The locus of all solutions as a function of the normalized central energy density $\Psi(0)$ and regeneration parameter $r\Sigma^2$. The different curves correspond to different fault half-widths w , in units of the scattering length. For fault widths larger than 1 nearly all solutions lie on similar curves in this plot. The minimum value of $r\Sigma^2$ for a solution to exist is about 2.8 if the width is greater than 1. The fault half-width w in this figure is in units of the scattering length l_s .

must be correspondingly large. Thus, we must have $r > 2.8/\Sigma^2$, or $r > 280$.

Assuming a seismic efficiency, e , of about²⁸ 0.5, this means that the intrinsic Q must be more than 1,000 at high frequencies. Seismic Q s are typically around 500 for the crust and upper mantle³⁵. However, the seismic Q includes both loss by conversion to heat and loss by scattering. In the past several years attempts have been made to estimate the intrinsic Q by accounting for scattered energy in the seismic coda. These estimates yield values of intrinsic Q in the range of 1,500 to 2,000 or more for the Earth's crust^{32,36}. Admittedly, these measurements are at frequencies only up to about 30 Hz, but they strongly suggest that the necessary high values of Q are realized in the Earth. Even if the intrinsic Q were as low as 100, it would still permit shearing flow at a substantially reduced coefficient of friction. Because the threshold stress Σ depends on the square root of e and Q it is not very sensitive to errors in estimates of their size. If e were as small as 0.1, this would raise the threshold stress by a factor of 2.3, still resulting in a substantially reduced coefficient of friction. Acoustic fluidization is ineffective only if the product eQ is as small as 1.

It is more difficult to estimate the velocity of flow because it is controlled by parameters that are even more poorly constrained than Q . In Fig. 2 the dimensionless velocity across the fault zone is about 1.4 (twice the value plotted for half the fault zone). Equation (8) must be used to convert this value to a real velocity, which means that values for scattering length and wavelength are needed. As wavelengths are short, and scattering is presumed to be strong, to a first approximation it can be assumed that the scattering length l_s is comparable to the wavelength λ . In this case the ratio is near 1, and

$$\dot{u} \approx \frac{\tau}{\rho c} v \quad (9)$$

This equation has precisely the same form as that derived by Brune³⁷ from elastodynamic considerations. Brune's equation has been shown to provide a good description of the particle velocity in the vicinity of an earthquake fault²⁸. Substituting values of $\tau = 10$ MPa, $\rho = 3,000$ kg m⁻³, and $c = 4$ km s⁻¹ gives $\dot{u} = 1.2$ m sec⁻¹ which is fairly typical for observed earthquakes.

Although slip occurs at close to the observed rate, this does not imply an extremely low viscosity in the fault zone. For strong acoustic vibrations the effective viscosity is of the order of $\eta = \rho\lambda c$. Assuming that most of the shear is confined to a zone of about

10 cm width near the core of the fault zone and that λ is about 10 cm also, η is about 10^6 Pa s, about the viscosity of very viscous, silica-rich lava. In this case we would not expect the gouge in the fault core to commonly develop injection structures, nor are such injection structures commonly observed²⁸.

The steady-state solutions described in Fig. 3 are converted to shear stress and differential slip velocity in Fig. 4 using equation (9) and the above parameters for velocity. It is easy to see that the upper branch corresponds to high-velocity slip at stresses well below the value given by a static coefficient of friction (vertical arrow on lower axis). These solutions are candidates for seismic slip on faults.

The lower branch also represents slip at low shear stress, but the velocities are much lower, only centimetres per second. Moreover, these solutions have the peculiar property that an increase of shear stress leads to a *decrease* in velocity. This kind of behaviour has been seen before in geophysics, in the motion of plates controlled by thermal runaway³⁹. Such solutions are not stable for fixed shear stress boundary conditions: small perturbations in the energy density grow rapidly and force the solution toward the upper branch, where the velocity increases rapidly and the result is an 'earthquake'. However, if different boundary conditions are chosen, such as constant slip velocity, then these solutions may be stable. Which type of boundary conditions are realized on an actual fault zone is difficult to say, but the very rarity of either fault creep⁴⁰ or silent earthquakes suggests that conditions stabilizing the lower solution branch may be very rare, assuming that these solutions really represent such phenomena.

Although the lower branch of solutions represents flow at low velocity, the computed velocities of a few centimetres per second are still far higher than has been observed on the creeping section of the San Andreas fault, where maximum velocities of about 10^{-6} m s⁻¹ are observed in hours-long "strain pulses"⁴¹. (Note that there is some controversy about whether such strain pulses are deep-seated or shallow soil phenomena; the process of fault creep is not yet well understood.) But before discarding this explanation for fault creep, note that because the elastic energy densities are so low for this class of solutions, they are very sensitive to the assumption, equation (4), that the energy amplitude distribution follows a gaussian distribution, because the low energy densities pushes the solution far out onto the wings of the probability distribution. Another distribution would thus lead to quantitatively different results, although the general structure of the solution would be the same. In particular, it may be that the velocity is quite different than given by the gaussian distribution. The upper branch of solutions is not subject to this proviso, because the amplitude of the elastic waves is much larger and the pressure fluctuations are close to the centre of the probability distribution.

On the other hand, the low-velocity branch may correspond to the type of movement that has been observed in "slow" or "silent" earthquakes⁴². Such events seem to be most common on oceanic transform faults where the boundary conditions may differ from those of other faults. Whatever these new solutions correspond to, it is interesting that the acoustic fluidization mechanism gives rise to two distinct modes of fault slip, a 'fast' mode and a 'slow' mode which coincide only at the minimum stress threshold for slip.

Fault strength and earthquake initiation

The process of acoustic fluidization offers a mechanism by which strong vibrations facilitate the flow of incoherent rock debris in the cores of mature faults. However, some faults appear to be strong¹⁰: what determines whether a fault is strong or weak? The mechanism described assumes that an incohesive fault core exists that can flow readily when the overburden is relieved. The fault surface of immature faults, however, is characterized by overlapping and interlocking fractures that inhibit ready deformation even without an overburden stress⁴³; the fault has an effective cohesion due to this block interaction^{44,45}. A fault develops an incohesive zone near

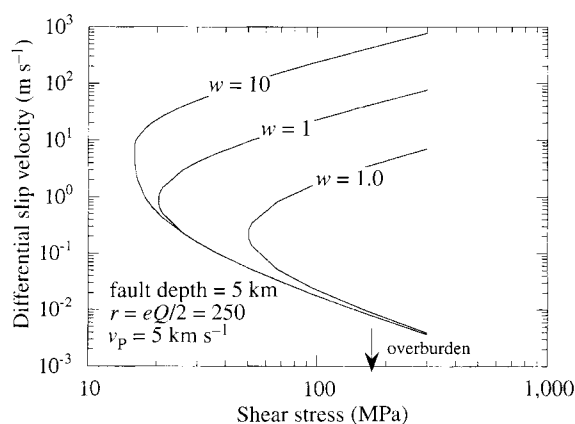


FIG. 4 The same solutions in Fig. 3, but displayed as a function of driving stress shear and differential slip velocity for plausible values of the material parameters (see text). The vertical arrow indicates the overburden stress at the mean depth (5 km) of the fault. Slip can clearly proceed under stresses much lower than the overburden, at values comparable to the observed estimates for regional stress driving the San Andreas fault. The fault width w in this figure is in units of the scattering length l_s . The quantity r is the regeneration parameter, defined in terms of the seismic efficiency e and inverse dissipation Q . v_p is the P-wave velocity.

its core only after substantial slip has accumulated³⁸. Furthermore, the lack of scattering inhomogeneities near the plane of fresh faults may permit the escape of too much vibrational energy from the vicinity of the fault plane, whereas in older faults a metres-wide fracture zone²⁰ may scatter seismic energy more effectively back to the fault core.

This Article has only addressed the properties of steady-state solutions to the simplified one-dimensional acoustic fluidization equations. However, an important question is how does the process get started. Clearly, if there is no acoustic energy to begin with, nothing can happen. The minimum requirement for the development of a non-trivial steady-state solution (an earthquake) is that the regional stress exceeds the threshold given above. In addition, some other event must start the process by generating enough acoustic energy for regeneration to make up for losses due to scattering and absorption. Thus some kind of trigger event, perhaps a high stress-drop event in a limited area of the fault, must generate enough vibration over a large enough volume to permit the acoustic energy field to sustain itself and grow. The necessity of a 'sufficiently large volume' may be why this process has not been observed in the laboratory: the small scale of laboratory studies means that surface area is large compared to the volume, thus enhancing the loss of acoustic energy through radiation. Study of this type of time- and volume-dependent behaviour is a project for future work.

Proof that acoustic fluidization really does account for the low strength of mature faults will not be easy to obtain. Because the strong, high-frequency vibrations are attenuated rapidly away from the fault plane they probably cannot be detected except by *in situ* measurements. What is required is thus one (or, ideally, several) velocity gauges emplaced less than a metre from the fault core. During an earthquake (assuming the gauges continue to function) they should record vibrations with amplitudes comparable to the overburden pressure and frequencies in the vicinity of a few kHz. Although such measurements may be very ambitious, projects such as the Cajon Pass borehole into the San Andreas fault zone may someday make them possible. In the meantime, laboratory experiments on vibrated granular materials may be able to check such things as the validity of equation (1) or the amplitude and frequency of sound generated during shearing. □

Received 18 August 1995; accepted 17 January 1996.

1. Henyey, T. L. & Wasserburg, G. J. *J. geophys. Res.* **76**, 7924–7945 (1971).
2. Brune, J. N., Henyey, T. L. & Roy, R. F. *J. geophys. Res.* **74**, 3821–3827 (1969).
3. Lachenbruch, A. H. & Sass, J. H. *J. geophys. Res.* **97**, 4995–5015 (1992).
4. Lachenbruch, A. H. & Sass, J. H. *J. geophys. Res.* **85**, 6185–6222 (1980).
5. Zoback, M. D. & Healy, J. H. *J. geophys. Res.* **97**, 5039–5057 (1992).
6. Zoback, M. D. & Roller, J. C. *Science* **206**, 445–447 (1979).
7. Mount, V. S. & Suppe, J. *Geology* **15**, 1143–1146 (1987).
8. Zoback, M. D. et al. *Science* **238**, 1105–1111 (1987).
9. Hickman, S. H. *Rev. Geophys. (suppl.)* 759–775 (1991).
10. Sibson, R. H. *Pure appl. Geophys.* **142**, 645–662 (1994).
11. Sleep, N. H. & Blanpied, M. L. *Nature* **359**, 687–692 (1992).
12. Byerlee, J. *Geology* **21**, 303–306 (1993).
13. Rice, J. R. in *Fault Mechanics and Transport Properties in Rocks* (eds Evans, B. & Wong, T.-F.) 475–503 (Academic, San Diego, 1992).
14. Lockner, D. A. & Byerlee, J. D. *Geophys. Res. Lett.* **21**, 2353–2356 (1994).
15. Melosh, H. J. *J. geophys. Res.* **84**, 7513–7520 (1979).
16. Brune, J. N., Brown, S. & Johnson, P. A. *Tectonophysics* **218**, 59–67 (1993).
17. Oden, J. T. & Martins, J. A. C. *Comp. Meth. appl. Mech. Engng.* **52**, 527–634 (1985).
18. Heaton, T. H. *Phys. Earth planet. Inter.* **64**, 1–20 (1990).
19. Foda, M. A. *Proc. R. Soc. Lond. A* **435**, 233–244 (1991).
20. Chester, F. M., Evans, J. P. & Biegel, R. L. *J. geophys. Res.* **98**, 771–786 (1993).
21. Byerlee, J. *Pure appl. Geophys.* **116**, 615–626 (1978).
22. Crandall, S. H. & Mark, W. D. *Random Vibration in Mechanical Systems* (Academic, New York, 1973).
23. Rodger, A. A. & Littlejohn, G. S. *Géotechnique* **30**, 269–293 (1980).
24. Richards, R., Elms, D. G. & Budhu, M. *J. Geotech. Engng.* **116**, 740–759 (1990).
25. Barkan, D. D. *Dynamics of Bases and Foundations* (McGraw-Hill, New York, 1962).
26. Zik, O., Stavans, J. & Rabin, Y. *Europhys. Lett.* **17**, 315–319 (1992).
27. Blekhnman, I. I. in *Advances in Theoretical and Applied Mechanics* (eds Ishlinsky, A. & Chermousko, F.) 127–147 (MIR, Moscow, 1981).
28. Kanamori, H. A. *Rev. Earth planet. Sci.* **22**, 207–237 (1994).
29. Dainty, A. M., Toksöz, M. N., Anderson, K. R. & Pines, P. S. *The Moon* **2**, 11–29 (1974).
30. Dainty, A. M. & Toksöz, M. N. *J. Geophys.* **43**, 375–388 (1977).
31. Nakamura, Y. *J. Geophys.* **43**, 389–399 (1977).
32. Frankel, A. & Wennerberg, L. *Bull. seism. Soc. Am.* **77**, 1223–1251 (1987).
33. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ. Press, 1986).
34. Bergé, P., Pomeau, Y. & Vidal, C. *Order within Chaos* (Wiley, New York, 1984).
35. Anderson, D. L. *Theory of the Earth* (Blackwell, Boston, 1989).
36. Dainty, A. M. *Geophys. Res. Lett.* **8**, 1126–1128 (1981).
37. Brune, J. N. *J. geophys. Res.* **75**, 4997–5009 (1970).
38. Engelder, J. T. *Geol. Soc. Am. Bull.* **85**, 1515–1522 (1974).
39. Melosh, H. J. & Ebel, J. *Geophys. J. R. astr. Soc.* **59**, 419–436 (1979).
40. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
41. Wesson, R. L. *J. geophys. Res.* **93**, 8929–8951 (1988).
42. Beroza, G. C. & Jordan, T. H. *J. geophys. Res.* **95**, 2485–2510 (1990).
43. Sibson, R. H. *Pure appl. Geophys.* **124**, 159–175 (1986).
44. Goodman, R. A., Taylor, R. L. & Brekke, T. L. *J. Soil Mech. Fdn. Div. Am. Soc. civ. Engrs* **94**, 637–659 (1968).
45. Jaeger, J. C. *Géotechnique* **21**, 97–134 (1971).

ACKNOWLEDGEMENTS. This work was supported in part by the US NSF.

Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter

Bruno Giros*†, Mohamed Jaber*, Sara R. Jones*, R. Mark Wightman‡ & Marc G. Caron*

* Howard Hughes Medical Institute Laboratories, Departments of Cell Biology and Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA

‡ Curriculum in Neurobiology and Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27514, USA

Disruption of the mouse dopamine transporter gene results in spontaneous hyperlocomotion despite major adaptive changes, such as decreases in neurotransmitter and receptor levels. In homozygote mice, dopamine persists at least 100 times longer in the extracellular space, explaining the biochemical basis of the hyperdopaminergic phenotype and demonstrating the critical role of the transporter in regulating neurotransmission. The dopamine transporter is an obligatory target of cocaine and amphetamine, as these psychostimulants have no effect on locomotor activity or dopamine release and uptake in mice lacking the transporter.

THE dopamine transporter (DAT) is believed to control the temporal and spatial activity of released dopamine (DA) by rapid uptake of the neurotransmitter into presynaptic terminals. It is therefore an important element in regulating the action of DA on locomotion, cognition, affect and neuroendocrine functions^{1–3}. The DAT is also a target for psychoactive drugs, such as antidepressants and drugs of abuse including cocaine and amphetamine^{4,5}, and is known to act as a gate for several neurotoxins that destroy DA neurons (notably 6-hydroxydopamine, 1,2,3,6-tetrahydro-1-methyl-4-phenylpyridine (MPTP) and methamphetamine⁶). Elucidation of the primary structure of the DAT has revealed that it is a member of the large family of Na⁺/Cl[−]-dependent transporters containing 12 putative transmembrane domains^{2,3}. Other members of this family include transporters for serotonin, norepinephrine, GABA (γ-aminobutyric acid), proline, glycine, creatine, taurine and betaine^{2,3}. Members of the trans-

porter family display substrate specificity, although a small degree of overlap exists amongst some of the members. The use of transporter mutants and chimaeric proteins has identified residues and domains of the DA, norepinephrine and serotonin transporters involved in substrate specificity and stereoselectivity, as well as antidepressant/psychostimulant interactions^{7–10}. Monoamine transporters are each the product of a single gene². Studies of their distribution have shown that their expression is restricted to cells that synthesize the corresponding neurotransmitter, thus providing anatomical segregation that contributes to the specificity of the transport process. Accordingly, the DAT messenger RNA is highly expressed in dopaminergic cells of the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA)^{11,12}, and the transporter protein is found in abundance in dopaminergic projection areas of the basal ganglia¹³.

Neurotransmission is a complex process. It requires synthesis and storage of the neurotransmitter, followed by its release and interaction with specific pre- and postsynaptic receptors.

† Present address: INSERM U-288, 91, Boulevard de l'Hôpital, 75013 Paris, France.

Termination of the signal involves enzymatic degradation of the neurotransmitter, its diffusion out of the synapse, and/or its removal from the synapse by one of the Na^+/Cl^- -dependent neurotransmitter transporters. The importance of each of these processes for DA neurotransmission has been previously deduced, mainly from pharmacological manipulations. However, several of these processes have since been investigated directly by targeted disruption of the genes coding for tyrosine hydroxylase¹⁴, as well as D1 and D2 dopamine receptors^{15–17}. A direct assessment of the role of transporters in the development, function and maintenance of a neurotransmitter system has not yet been possible, although the importance of transporters in neuronal activity has long been appreciated from the therapeutic action of classical uptake blockers such as the antidepressants¹⁸.

As well as removing extracellular transmitters, monoamine transporters have been proposed as targets of various drugs of abuse. In particular, the DAT has been suggested to mediate the reinforcing/addictive properties of cocaine and amphetamine as these effects can be attenuated by selective dopamine receptor antagonists¹⁹. However, the molecular mechanisms by which these drugs exert their effects are different. Cocaine, like most compounds that interact with transporters, is a classical uptake blocker; it inhibits DA uptake after normal neuronal activity⁴. Amphetamine-like drugs, however, are thought to have a more complex mechanism of action, promoting the actual release of the transmitter in addition to inhibiting uptake and mediating other intracellular effects^{20–23}. Whether all actions of psychostimulants on the DA system are strictly DAT dependent has remained controversial.

Here we report the disruption of the mouse DAT gene by homologous recombination²⁴. Our results establish not only the central importance of the transporter as the key element controlling synaptic dopamine levels, but also its role as an obligatory target for the behavioural and biochemical action of amphetamine and cocaine.

Disruption of the dopamine transporter

We have produced, by means of *in vivo* homologous recombination, a strain of mice in which the gene encoding the DAT has been disrupted (Fig. 1a). The homozygote ($\text{DAT}^{-/-}$) mice showed the expected pattern of gene disruption, as established by Southern blotting (Fig. 1b). No mRNA encoding the DAT was detected in the VTA or the SNC of homozygote animals, as demonstrated by *in situ* hybridization (Fig. 1c). Striatal synaptosomal preparations from $\text{DAT}^{-/-}$ mice showed no significant DA uptake (Fig. 1d), suggesting that other monoamine transporters do not contribute to DA uptake.

Homozygote mice gain weight more slowly than heterozygote and wild-type ($\text{DAT}^{+/+}$) mice (Fig. 2a) and show a significant propensity to premature death (68% survival after 10 weeks in homozygotes, as compared to 97% in the wild type) (Fig. 2b). These observations may be linked to impaired food intake, as blockade of DAT with cocaine is associated with modification of dietary habits²⁵. $\text{DAT}^{-/-}$ mice are fertile and mate, but females have impaired maternal behaviour, so offspring have to be transferred to foster mothers to be successfully raised to adulthood. This observation is consistent with the known role of DA on pituitary gland function, and supports the controversial notion that the DAT participates in the regulation of the tuberoinfundibular DA pathway^{26,27}. In addition, general modifications of cognitive or affective functions might also underlie this phenotype. Previous reports indicate that cocaine treatment perturbs maternal behaviour in rats²⁸.

Locomotor activity

The DAT plays a major role in the control of locomotor behaviour by regulating dopaminergic tone in the basal ganglia and nucleus accumbens. Drugs that interfere with DA uptake, such as the psychostimulants cocaine and amphetamine, increase stereotypy and locomotor behaviour^{4,29–31}. The spontaneous locomotor

activity of naive homozygotes (those that were not habituated to the test) is highly elevated (Fig. 3a). They have a half-time of habituation twice as long (90 min compared to 40 min) as the wild-type or heterozygote mice, and are five to six times more active than wild-type animals during both phases of the light–dark cycle (Fig. 3b). Neither the heterozygotes nor the homozygotes exhibit a significantly enhanced level of spontaneous stereotypies (sniffing, grooming or rearing).

Because homozygotes lack one of the major targets of psychostimulant drugs, we investigated the behavioural effects of treatment with either cocaine or amphetamine. We found that intraperitoneal injections of high doses of cocaine (40 mg per kg) and D-amphetamine (10 mg per kg) produced no significant locomotor effects (Fig. 3c) or stereotyped behaviours (data not shown) in $\text{DAT}^{-/-}$ mice. In wild-type and heterozygote mice, treatment with either of these two psychostimulants produced a six- to eightfold increase in their locomotor activity, which was elevated by the same amount as the spontaneous activity in the homozygote mice. At these concentrations of psychostimulants, wild-type mice developed stereotyped behaviours with time. The lack of locomotor effects of psychostimulants on homozygote mice is not likely to be due to the homozygotes already being maximally hyperactive, and therefore unable to show an increase in their locomotor activity, because our studies were performed during the light cycle, when locomotor activity was significantly lower than during the dark cycle (Fig. 3b). These findings demonstrate directly that the DAT, and not other monoamine transporters, plays a crucial role in the hyperlocomotor effects of psychostimulants.

Dopaminergic regulation of gene expression

Dopamine exerts a modulatory control on dopaminergic and dopaminoceptive cells at the pre- and postsynaptic level³². In the basal ganglia, preproenkephalin A (PPA) is under the inhibitory influence of DA by means of the D2 receptors^{32,33}, whereas substance P (SP) and dynorphin (Dyn) are under the excitatory influence of DA by means of the D1 receptor^{32,34}. By *in situ* hybridization (Fig. 4a), PPA mRNA levels were reduced substantially (66%) and mRNA coding for Dyn was increased by 50%, but SP mRNA levels were not significantly modified in homozygote mice. Different activation or regulatory pathways may exist for Dyn and SP gene expression as their mRNAs were not both upregulated. For the two major DA receptors (D1 and D2), the coding mRNA was downregulated by 55% and 45%, respectively. Simultaneous decreases in D1 and D2 receptor mRNAs is unprecedented as, to the best of our knowledge, neither lesions nor pharmacological or behavioural manipulations of DA transmission have ever produced such a marked downregulation of both mRNAs. A putative downregulation of D2 autoreceptors, which are involved in the regulation of DA synthesis and release³⁵, was indicated by a 50% decrease in D2 receptor mRNA in DA cells (SNC and VTA). The mRNA changes described here reflect increased dopaminergic neurotransmission, which correlates well with the increase in spontaneous locomotor activity. Moreover, the findings that the heterozygotes consistently displayed intermediate values reinforces the idea that the DAT is important for the maintenance of normal dopaminergic neurotransmission, such that a 50% decrease in DAT levels produces profound changes in dopaminergic pathways (Fig. 4b).

Dopamine dynamics and psychostimulants

The consequences of the absence of the DAT on striatal DA neurotransmission were investigated in an attempt to determine the biochemical basis of the marked hyperlocomotor activity even in the presence of extensive adaptive changes. To achieve this we measured real-time transmitter dynamics using fast-scan cyclic voltammetry, an electrochemical detection technique based on the oxidation of DA, with a time resolution of 100 ms (ref. 36). This technique is specific in that the cyclic voltammograms generated provide unique fingerprints of the type of electroactive

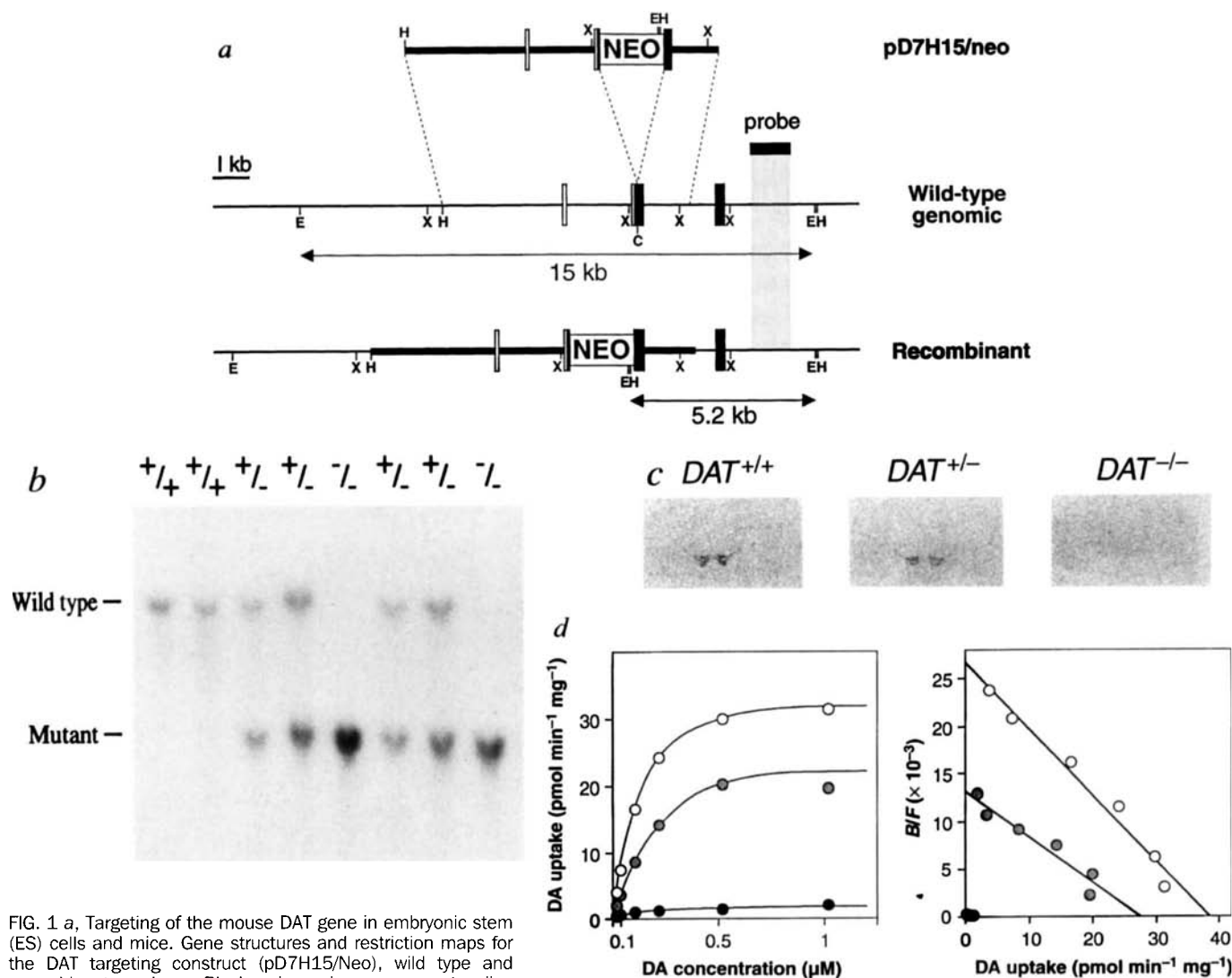


FIG. 1 **a**, Targeting of the mouse DAT gene in embryonic stem (ES) cells and mice. Gene structures and restriction maps for the DAT targeting construct (pD7H15/Neo), wild type and recombinant are shown. Black and open boxes represent coding and untranslated exons, respectively. Arrow lines indicate the EcoRI restriction fragment sizes for wild-type and recombinant DNA. Restriction sites are: C, *ClaI*; E, *EcoRI*; H, *HindIII*; X, *XbaI*. The probe used for Southern analysis is shown as a thick black line. **b**, Genomic Southern blot analyses of tail biopsies. Genomic DNA from offspring obtained following breeding of heterozygote mice was digested with *EcoRI* and probed with a 3' 1.1-kb genomic fragment. Wild-type and recombinant restriction fragments are 15 kb and 5.2 kb, respectively. **c**, *In situ* hybridization in ventral midbrain of wild-type ($DAT^{+/+}$), heterozygote ($DAT^{+/-}$) and homozygote ($DAT^{-/-}$) mice using three oligonucleotides derived from different parts of the DAT mRNA sequence⁴⁷. **d**, Uptake of [³H]dopamine into striatal synaptosomes of wild-type (open circles), $DAT^{-/-}$ (grey circles) and $DAT^{+/-}$ (black circles) mice. Dopamine saturation curve and Scatchard analysis are shown. Mean values $\pm$ s.e.m. (3 animals per group) for K_M (nM) and V_{max} (pmol per min per mg protein) are, respectively: wild type, 138 ± 22 , 38 ± 7 ; $DAT^{+/-}$, 199 ± 24 , 27 ± 4 ; $DAT^{-/-}$, 551 ± 78 , 1.5 ± 0.8 . β represents the amount of DA taken up during the incubation, and F the amount of free DA remaining.

METHODS. A rat DAT cDNA probe comprising the N-terminal to the third transmembrane domain was used to screen a mouse 129 Sv/J genomic library as described¹¹. A 12-kb phage containing the first three exons of the mouse DAT gene was isolated, and a 7.5-kb *HindIII* fragment (from an internal site to a phage cloning site) was excised and subcloned into pGEM4Z. An *EcoRI*–*SaI*–*NotI* adapter was introduced into a unique *ClaI* site at position 109 of the reading frame, and a *NotI* cassette containing the neomycin resistance (*neo*) gene under the control of the PGK promoter was introduced into the *NotI* site of the construct to provide pD7H15/neo. The *neo* cassette was cloned in the opposite direction to the DAT gene, and DNA sequencing was used to insure that pD7H15/neo did not contain a consensus start site that could be used to resume DAT translation. pD7H15/neo was linearized with *AatII* before electroporation into ES cells. Cell culture, electroporation of E14TG2a ES cells²⁴ and generation of chimaeric mice were performed essentially as previously described²⁴. For Southern analysis, 5–15 μg DNA from ES cells or tail biopsies were digested

with *EcoRI* or *HindIII* and probed with a 1.1-kb genomic DNA probe, and a *neo* probe to ensure that there was only one copy of the construct in ES cells (not shown). About 20% of the tested ES cells were positive for homologous recombination, and three clones were selected for karyotyping and injection into C57/B6J blastocysts. Chimaeric males were mated with C57/B6J females, and agouti coat pups were typed by Southern blotting of tail biopsies. Two germline transmitting males (from different ES clones) were used as colony founders. For *in situ* hybridization, mice were killed by decapitation and the brains were dissected out, immersed overnight in 1% paraformaldehyde, and cryoprotected in 15% sucrose–PBS buffer for 6–8 h. The brains were then frozen in liquid nitrogen and cut into frontal sections (12 μm) which were stored at -80°C until required. The *in situ* hybridization procedure was performed with oligonucleotide probes, labelled by tailing with [³⁵S]dATP (NEN), with a specific activity of 2×10^9 cpm μg⁻¹ as previously described^{31,47}. Sections were exposed at room temperature to X-ray films (Kodak X-Omat) for 7 days. Autoradiograms were scanned (Howtek Scanner) at 1200 d.p.i. and analysed (NIH-Image). Probe concentration and exposure times were chosen to stay within a linear range of detection. Experiments were performed in duplicate, and comparisons were made between groups using the Student's *t*-test with the control value ($DAT^{+/+}$) set equal to 1, and $DAT^{+/-}$ and $DAT^{-/-}$ values expressed relatively to it. For DA uptake experiments, two striata from one mouse were homogenized in 0.5 ml of 0.3 M sucrose and synaptosomes were prepared as described⁶. The synaptosomal preparation (20–25 μg protein per tube) was incubated with 0.25 μCi of [³H]dopamine (51 Ci mmol⁻¹), and increasing concentrations of DA (0–3 μM) for 5 min at 37°C in a final volume of 0.5 ml uptake buffer (containing, in mM, 4 Tris-HCl, 6.25 HEPES, 120 NaCl, 5 KCl, 1.2 CaCl₂, 1.2 MgSO₄, 5.6 D-glucose, 0.5 ascorbic acid, pH 7.1). Experiments were done in triplicate for each concentration, and nonspecific uptake was determined in the presence of 30 μM nomifensine. Uptake was stopped and counted as described⁶. K_M and V_{max} were calculated by the iterative curve-fitting programs EBDA and LIGAND⁶, and resulting values of three separate mice per group were averaged.

FIG. 2 *a*, Comparative postnatal weights of wild-type ($DAT^{+/+}$), heterozygote ($DAT^{+/-}$) and homozygote ($DAT^{-/-}$) mice. Weights (mean $\pm$ s.e.m.) of $DAT^{+/+}$ ($n = 63$, open circles), $DAT^{+/-}$ ($n = 103$, grey circles) and $DAT^{-/-}$ ($n = 59$, black circles) were recorded at birth, and every following week as indicated. *, $P < 0.05$ as compared to $DAT^{+/+}$, using Student's *t*-test; s.e.m. values are less than 5% of the mean unless stated otherwise. *b*, Percentage of postnatal survival of $DAT^{+/+}$ (open circles), $DAT^{+/-}$ (grey circles) and $DAT^{-/-}$ (black circles) mice. At day 1, there were 63, 103 and 59 animals for each group, respectively. From 333 offspring from the mating of heterozygotes, we obtain 78 $DAT^{+/+}$, 179 $DAT^{+/-}$ and 76 $DAT^{-/-}$, numbers very close to the ratio (1:2:1) which may be expected in the absence of prenatal death.

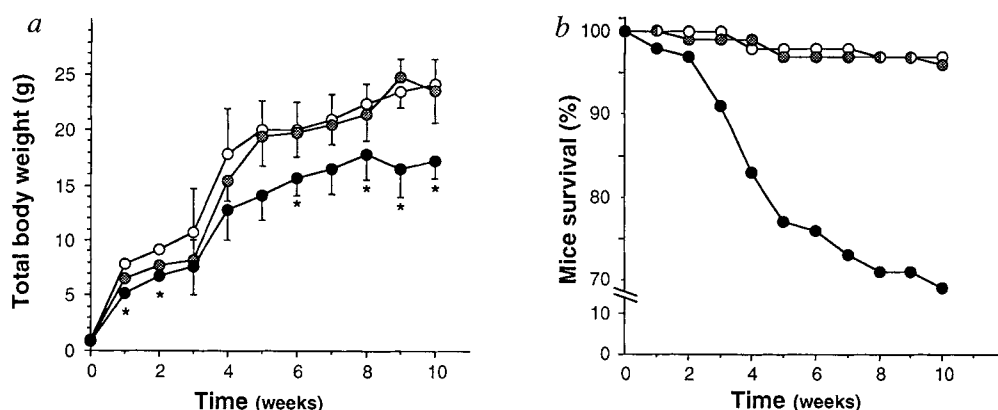
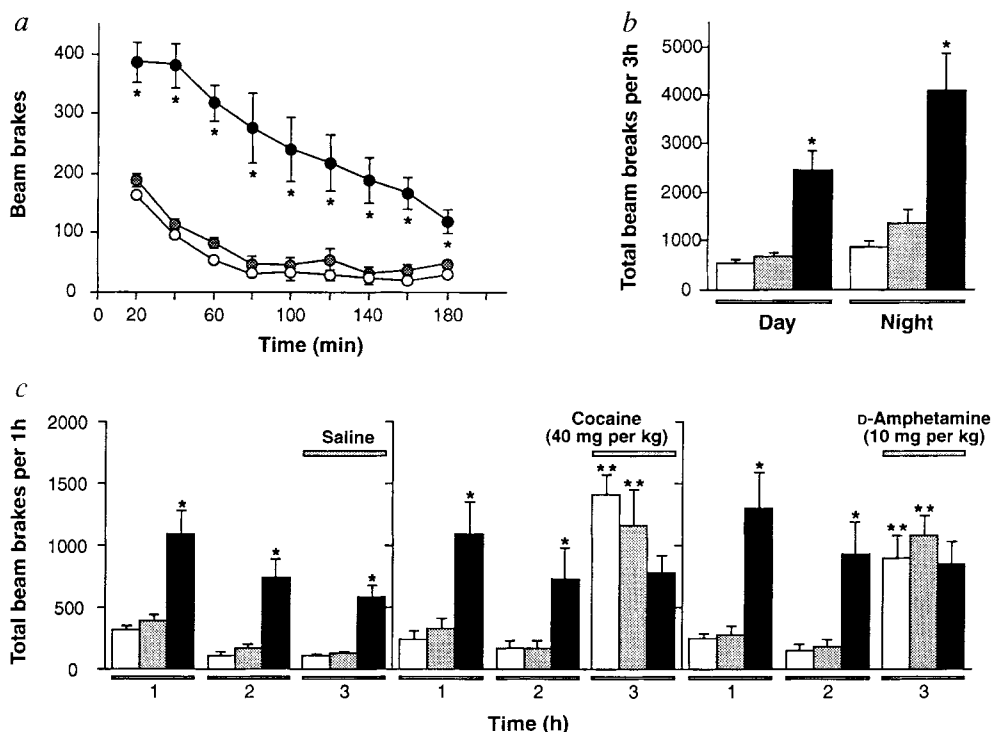


FIG. 3 *a*, Spontaneous locomotor activity and habituation of naive wild-type ($DAT^{+/+}$) (open circles), heterozygote $DAT^{+/-}$ (grey circles) and homozygote ($DAT^{-/-}$) (black circles) mice. Locomotor activity was recorded every 20 min for a period of 3 h, $n = 12$ mice per group. *, $P < 0.01$ as compared to $DAT^{+/+}$, using the Student's *t*-test; s.e.m. is less than 5% of the mean if not stated otherwise. *b*, Spontaneous locomotor activity of naive $DAT^{+/+}$ (open bars), $DAT^{+/-}$ (grey bars) and $DAT^{-/-}$ (black bars) mice. Accumulated locomotor activity was recorded for 3 h during the day (11:00–14:00) or the night (23:00–2:00) phase of the light–dark cycle. *, $P < 0.001$ as compared to $DAT^{+/+}$; $n = 10$ –12 mice per group. The spontaneous locomotor activity of the homozygote animals was significantly higher during the night cycle compared to the day cycle ($P < 0.05$). The heterozygotes are consistently more active than the wild-type mice but this increase is of marginal significance ($P = 0.06$) during the dark phase of the cycle. *c*, Effects of psychostimulants on the locomotor activity of $DAT^{+/+}$ (open bars), $DAT^{+/-}$ (grey bars) and $DAT^{-/-}$ (black bars) mice. Locomotor activity was recorded for 1 h time periods. After a 2-h habituation period, mice were injected with saline, 40 mg per kg cocaine or 10 mg per kg *D*-amphetamine, and locomotor activity was recorded for the next hour. *, $P < 0.001$ as compared to $DAT^{+/+}$; **, $P < 0.001$ as compared to the locomotor activity before injection; $n = 10$ –12 mice per group.



Locomotor activity was recorded for 1 h time periods. After a 2-h habituation period, mice were injected with saline, 40 mg per kg cocaine or 10 mg per kg *D*-amphetamine, and locomotor activity was recorded for the next hour. *, $P < 0.001$ as compared to $DAT^{+/+}$; **, $P < 0.001$ as compared to the locomotor activity before injection; $n = 10$ –12 mice per group. METHODS. Mice were maintained in standard housing conditions (12 h light–dark cycle) and were allowed free access to food and water. They were housed in groups of 6, and used at 6–10 weeks of age. Locomotor activity was measured in Plexiglas boxes (217 $\times$ 268 $\times$ 104 mm) with two photocell beams located across the long and the short axis, 15 mm above the

floor. Activity boxes were localized in dark cabinets (6 boxes per cabinet, 6 cabinets) in a quiet room, and the 3 groups of mice were analysed at the same time. Placement of the mice and recording of the activity numbers were done in double-blind fashion. Saline solution, cocaine (40 mg per kg) or *D*-amphetamine (10 mg per kg) were injected intraperitoneally. Statistical analyses were performed using the Student's *t*-test. Except when stated, experiments were performed during the light phase (11:00–16:00). Stereotyped behaviours were also assessed with wild-type and homozygote mice by the method of McLennan and Maier as described previously³¹.

substance detected³⁶. In addition, the time resolution allows actual visualization of release and uptake of the neurotransmitter³⁷. Changes in baseline levels of DA in mice of all genotypes were undetectable in the absence of pharmacological or electrical stimulations. Measurements of extracellular DA levels were made in striatal slices. Electrical stimulation was used to evoke calcium-dependent vesicular DA release, and the rate of DA release and uptake into presynaptic terminals was measured³⁷. The cyclic voltammograms (Fig. 5, inset) demonstrate that the electroactive species detected under all conditions was DA. As shown in Fig. 5, a single stimulus pulse applied to caudate slices from wild-type, heterozygote and homozygote DAT knockout mice evoked release of DA, which was subsequently removed from the extracellular space. The average concentration of released DA was $4.1 \pm 0.5 \mu\text{M}$ for the wild-type, $2.7 \pm 0.3 \mu\text{M}$ for the heterozygote, and $1.1 \pm 0.1 \mu\text{M}$ for the homozygote knockout mice. These results suggest that tissue DA content or releasable pools of DA are markedly decreased following inactivation of the DAT. Consistent with the reduced DA release, levels of the rate-limiting enzyme of DA synthesis, tyrosine hydroxylase, were decreased by approximately 90% in DA neurons and terminals in the homozygote animals (M. Jaber *et al.*, in preparation).

The time for DA clearance was approximately 1 s in wild-type and 100 s in homozygote mice. These results demonstrate that, in the absence of DAT, DA remains in the extracellular space at least 100 times longer than in the presence of a full complement of DAT. The timescale of DA clearance in homozygote animals is consistent with that predicted for diffusion in the slice³⁸. Thus the marked increase in spontaneous locomotor activity exhibited in homozygote mice is a direct consequence of the extended length of time that DA spends in the extracellular space following release. These results demonstrate that the DAT is the primary

mechanism for inactivation of DA, and indicate that the transporter is the most important determinant of transmitter signalling and tone.

The absence of enhanced locomotor activity and stereotypy after administration of amphetamine in the DAT knockout homozygote animals is intriguing. Amphetamine is known to release DA into the extracellular space, but the exact mechanism of DA transit across the plasma membrane is still controversial^{20,39,40}. Following incubation with $10 \mu\text{M}$ amphetamine, baseline non-stimulated release of DA was monitored by cyclic voltammetry in striatal slices from wild-type and homozygote mice. At this concentration amphetamine is known to enter neurons both by means of the DAT and by lipophilic diffusion^{20,39,41,42}. Amphetamine progressively increased levels of extracellular DA in wild-type mice (Fig. 6a), as identified by the cyclic voltammograms shown in Fig. 6b. However, extracellular levels of DA remained unchanged for at least 30 min in homozygote animals (Fig. 6a). In contrast to amphetamine, cocaine failed to induce spontaneous DA release in wild-type and homozygote slices (Fig. 6a), consistent with its proposed role as a DAT blocker that is dependent on neuronal activity. These results demonstrate that the main action of amphetamine is to reverse the transport process, and show that, in the absence of DAT, amphetamine does not lead to release of DA. Thus the failure of amphetamine to induce release of DA in homozygote striatal slices provides the biochemical explanation for its lack of effect on locomotor and stereotyped behaviours in the homozygote DAT knockout animals.

Discussion

Intense research and therapeutic interest have been directed at understanding DA neurotransmission because of its involvement in several normal, as well as dysfunctional, behaviours^{2,3}. Dop-

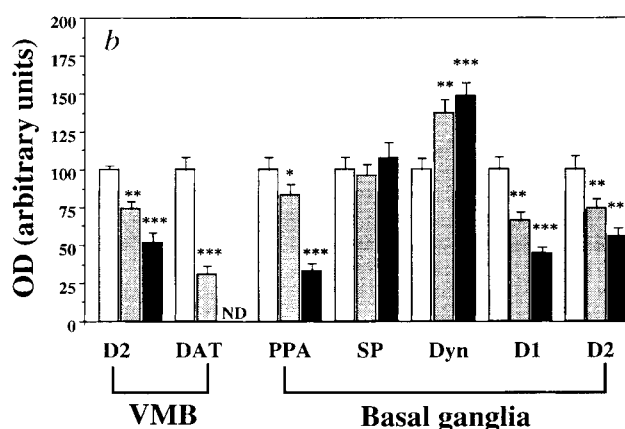
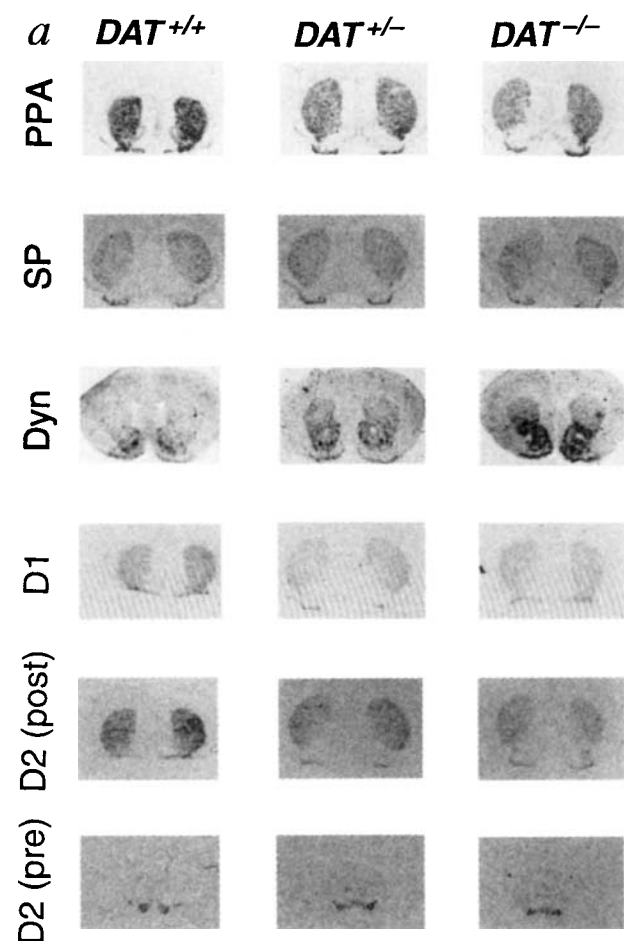


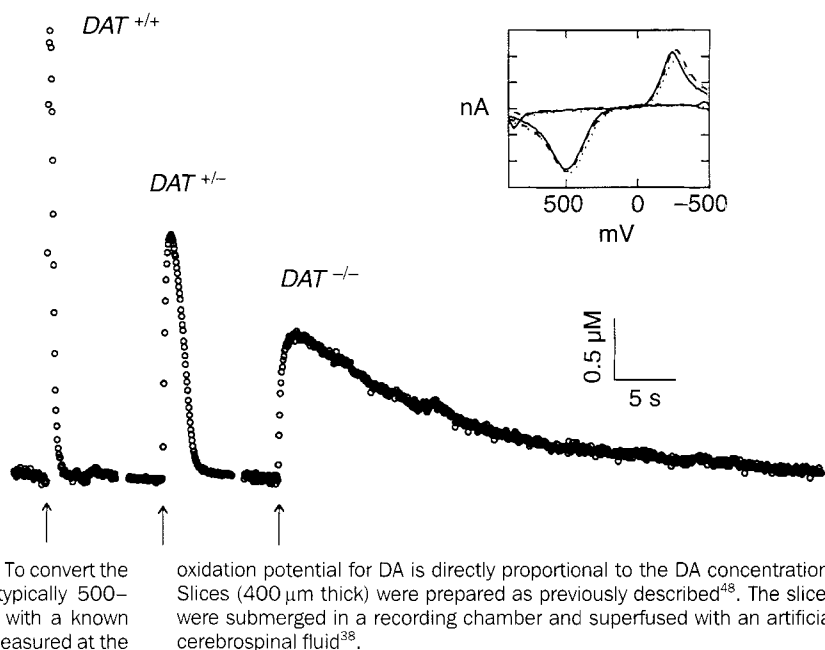
FIG. 4 *In situ* hybridization of various genes under dopaminergic control in the basal ganglia and the ventral midbrain of wild-type ($\text{DAT}^{+/+}$), heterozygote ($\text{DAT}^{+/-}$) and homozygote ($\text{DAT}^{-/-}$) mice. In the basal ganglia, detection was performed with specific antisense oligonucleotides recognizing SP, PPA, Dyn, D1 and D2 DA receptor mRNAs^{31,47}. In the ventral midbrain the D2 DA receptor mRNA (SNC and VTA) of $\text{DAT}^{+/+}$, $\text{DAT}^{+/-}$ and $\text{DAT}^{-/-}$ mice was also investigated. a, Autoradiograms of *in situ* hybridization in the basal ganglia and ventral midbrain; post, D2 postsynaptic receptor mRNA; pre, D2 presynaptic receptor mRNA. b, mRNA quantification of *in situ* hybridization autoradiograms. Images were analysed by the NIH image program and results expressed in arbitrary units of optical density (OD). VMB, ventral mid brain indicating VTA and SNC. Six mice per group were analysed (see Fig. 1 for methods), and experiments performed in duplicate or triplicate. Values were compared using the Student's *t*-test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ND, not detectable.

amine is involved in the control of motor function, cognition and affect¹. Imbalance in DA systems is thought to underlie conditions such as schizophrenia⁴³, Parkinson's disease⁴⁴, tardive dyskinesia⁴⁴ and drug addiction^{4,30,45}. Drugs of abuse such as cocaine and amphetamine target the dopaminergic system by increasing DA concentrations in the extracellular space and producing dramatic behavioural and biochemical changes^{1-3,21}.

Here we describe the first gene inactivation of a Na^+/Cl^- -dependent transporter. We demonstrate that the DAT is the most critical component in terminating DA neurotransmission, such that no other compensatory mechanism can overcome its long-term inactivation. Furthermore, we show that the DAT, in addition to its proposed role in the reinforcing properties of the psychostimulants cocaine and amphetamine⁴, represents an

FIG. 5 Typical recordings of electrically stimulated DA efflux in striatal slices from $\text{DAT}^{+/+}$, $\text{DAT}^{+/-}$ and $\text{DAT}^{-/-}$ mice. A single biphasic electrical pulse (4 ms, 350 μA) was applied at the time indicated by the arrows with a locally placed bipolar electrode⁴⁸. Data points (open circles) were collected from a recording electrode every 100 ms. Inset shows cyclic voltammograms recorded at the peak of DA efflux in each slice. Voltammograms recorded from $\text{DAT}^{+/+}$ (solid line), $\text{DAT}^{+/-}$ (broken line) and $\text{DAT}^{-/-}$ (dotted line) mice were scaled to approximately the same peak amplitude for comparison. These cyclic voltammograms are not significantly different from that obtained with authentic DA. Measurements were made in at least 22 slices from at least 5 animals for each group. The average time to clear released DA was 1, 3 and 100 s in $\text{DAT}^{+/+}$, $\text{DAT}^{+/-}$ and $\text{DAT}^{-/-}$, respectively.

METHODS. Carbon-fibre microelectrodes were prepared as described previously³⁶. A potentiostat (EI-400, Enscan Instrumentation, Bloomington, IN) was used for fast-scan cyclic voltammetry. The electrode potential was linearly scanned from -400 to 1,000 mV and back to -400 mV at 300 V s^{-1} every 100 ms. To convert the magnitude of current at the oxidation potential for DA (typically 500–700 mV) to concentration, each electrode was calibrated with a known concentration of DA at the end of the experiment. Current measured at the



oxidation potential for DA is directly proportional to the DA concentration. Slices (400 μm thick) were prepared as previously described⁴⁸. The slices were submerged in a recording chamber and superfused with an artificial cerebrospinal fluid³⁸.

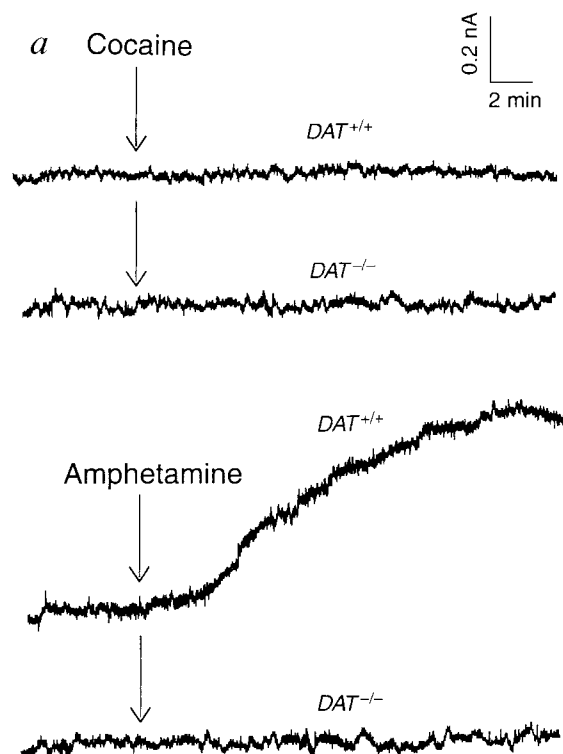
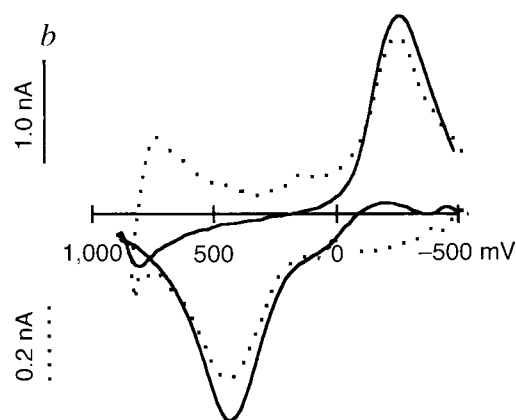


FIG. 6 a, Effect of 10 μM cocaine and 10 μM amphetamine on basal DA efflux in striatal slices from $\text{DAT}^{+/+}$ and $\text{DAT}^{-/-}$ mice. Drugs were applied at the times indicated by the arrows. Basal efflux of DA in the absence of electrical stimulation is shown over a 30-min period for each condition.



Data were collected every 100 ms for six 5-min periods and the points plotted as solid lines for clarity. The time before the arrows shows that there were no significant changes in baseline before drug application. Cocaine had no effect on basal DA efflux in $\text{DAT}^{+/+}$ or $\text{DAT}^{-/-}$ mice. Amphetamine caused basal DA efflux in the range of 3–10 μM in the $\text{DAT}^{+/+}$ mice, but had no effect on the baseline levels of DA in $\text{DAT}^{-/-}$ mice. At this concentration, amphetamine is known to enter the cell by lipophilic diffusion^{20,39,41,42}. Results are typical of at least 6 animals for each group. b, Background-subtracted cyclic voltammograms (current-voltage plot) recorded in a slice after amphetamine (broken line), overlaid with a voltammogram of authentic DA recorded in a calibration system (solid line). Note that the current axis is different for the two traces. The voltammograms are nearly identical, showing that DA is the primary substance being detected after amphetamine in a slice.

METHODS. Background-subtracted cyclic voltammograms were constructed by subtracting the current obtained after 5 min from that obtained 10 min after amphetamine application. Traces are the average of 10 voltammograms.

obligatory target for their locomotor as well as biochemical effects on the DA system.

Inactivation of the DAT in homozygote mice produces spontaneous hyperlocomotor activity which is virtually indistinguishable from that obtained with maximal doses of psychostimulants in wild-type animals (Fig. 3c). This phenotype is accompanied by profound alterations in the biochemical parameters of the nigrostriatal DA system, such as decreases in the mRNA and protein (data not shown) for D1 and D2 receptors, the main targets of synaptic DA responsiveness, as well as decreases in DA and tyrosine hydroxylase (data not shown), the rate-limiting enzyme of DA biosynthesis. These changes are of a magnitude not previously observed with any experimental or physiological conditions, and are all in a direction expected to blunt dopaminergic neurotransmission. Thus, given this profound downregulation, the spontaneous hyperlocomotor phenotype of homozygote mice was surprising and unexpected. However, real-time measurements of DA release and uptake show that, in the absence of the transporter, DA persists for a protracted period in the extracellular space, and that diffusion is the only remaining mechanism for clearance of DA. These findings provide the biochemical basis for the hyperactive phenotype and demonstrate that the DAT is the most crucial component regulating and maintaining dopaminergic transmission. Despite the marked changes in biochemical parameters, no obvious gross anatomical alterations in organization of the dopaminergic cell bodies or their projections were evident, which suggests that these structures develop normally (M. Jaber *et al.*, in preparation). It is not yet known whether the properties elucidated here for the disruption of the DAT will apply to other neurotransmitter transporters.

The lack of further locomotor activity or stereotypy in response to treatment of homozygote mice with high doses of cocaine or amphetamine suggests that the DAT is the major molecular target of these psychostimulants and mediates their behavioural and molecular action on DA transmission. This conclusion is further corroborated by real-time measurements using cyclic voltammetry of DA release and uptake in brain slices. Whereas cocaine and

amphetamine can inhibit the uptake of electrically stimulated release DA in wild-type and heterozygote mice, there was a total lack of effect of these drugs on the clearance of DA in homozygotes (data not shown). Using the same approach, we demonstrate that the main action of amphetamine is to release DA into the extracellular space in a strictly DAT-dependent manner. Many reports have already suggested that amphetamine might indeed exert its DA-releasing action by reversing the transporter, as classical uptake blockers can prevent amphetamine-mediated DA release^{20–23}. These studies, however, do not distinguish between the ability of the blockers to inhibit uptake of amphetamine into the neurons, actual release of DA, or any other possible membrane anaesthetic effects⁴⁶. As far as we are aware, this is the first direct and real-time demonstration that the DAT is a mandatory component for calcium-independent DA release triggered by amphetamine, to the exclusion of any other mechanism.

The DAT knockout mice should be an excellent tool for the study and development of drugs used in the management of dopaminergic dysfunction. There are similarities between some of the positive symptoms of schizophrenic patients⁴³ and the hyperdopaminergic phenotype of these knockout mice, so they provide a new animal model for the investigation of typical versus atypical neuroleptics. These mice will also aid the investigation of the role of dopaminergic neurotransmission in complex behavioural paradigms such as reward, addiction and tolerance properties of drugs of abuse. Although the evidence is strong that DA plays an important role in the rewarding effects of psychostimulants, as well as of opiates and ethanol, the exact nature of that role is not clear^{4,30}. Finally, these studies suggest exciting new avenues for the development of therapeutic strategies aimed at the treatment of neurological diseases. Indeed, the increased locomotor activity of DAT knockout mice, despite profound adaptive changes, suggests that specific blockade of the DAT with high-affinity inhibitors in illnesses such as Parkinson's disease, where the effective levels of DA are markedly reduced⁴⁴, may have beneficial consequences. □

Received 3 October 1995; accepted 11 January 1996.

- Iversen, L. L. *Br. J. Pharmac.* **41**, 571–591 (1971).
- Amara, S. G. & Kuhar, M. J. *A. Rev. Neurosci.* **16**, 16–73 (1993).
- Gros, B. & Caron, M. G. *Trends Pharmac. Sci.* **14**, 43–49 (1993).
- Ritz, M. C., Lamb, R. J., Goldberg, S. R. & Kuhar, M. J. *Science* **237**, 1219–1223 (1987).
- Horn, A. S. *Prog. Neurobiol.* **34**, 387–400 (1990).
- Pfiff, C., Gros, B. & Caron, M. G. *J. Neurosci.* **13**, 4246–4253 (1993).
- Kitayama, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 7782–7785 (1992).
- Gros, B. *et al. J. Biol. Chem.* **269**, 15985–15988 (1994).
- Buck, K. J. & Amara, S. G. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12584–12588 (1994).
- Barker, E. L., Kimmel, H. L. & Blakely, R. D. *Molec. Pharmac.* **46**, 799–807 (1994).
- Gros, B., El Mestikawy, S., Bertrand, L. & Caron, M. G. *FEBS Lett.* **295**, 149–154 (1991).
- Cerruti, C., Walther, D. M., Kuhar, M. J. & Uhl, G. R. *Molec. Brain Res.* **18**, 181–186 (1993).
- Ciliax, B. J. *et al. J. Neurosci.* **15**, 1714–1723 (1995).
- Zhou, Q.-Y., Quaipe, C. J. & Palmiter, R. D. *Nature* **374**, 640–643 (1995).
- Drago, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 12564–12568 (1994).
- Xu, M. *et al. Cell* **79**, 729–742 (1994).
- Balk, J.-H. *et al. Nature* **377**, 424–428 (1995).
- Fuller, R. W., Wong, D. T. & Robertson, D. W. *Med. Res. Rev.* **11**, 17–34 (1991).
- Wise, R. A. & Bozarth, M. A. *Psychiat. Med.* **3**, 445–460 (1985).
- Fisher, J. F. & Cho, A. K. *J. Pharmac. exp. Ther.* **208**, 203–209 (1979).
- Hitri, A., Hurd, Y. L., Wyatt, R. J. & Deutsch, S. I. *Clin. Neuropharmac.* **17**, 1–22 (1994).
- Eshelman, A. Y., Henningsen, R. A., Neve, K. A. & Janowsky, A. J. *Molec. Pharmac.* **45**, 312–316 (1994).
- Sulzer, D. *et al. J. Neurosci.* **15**, 4102–4108 (1995).
- Kohler, B. H. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8932–8935 (1989).
- Byck, R. & VanDyke, C. *NIDA Monogr.* **13**, 97–118 (1977).
- Meister, B. & Elde, R. *Neuroendocrinology* **58**, 388–395 (1993).
- Baumann, M. H. & Rothman, R. B. *Brain Res.* **608**, 175–179 (1993).
- Zimmerberg, B. & Gray, M. S. *Physiol. Behav.* **52**, 379–384 (1992).
- Kelly, P. H., Sevrour, P. W. & Iversen, S. D. *Brain Res.* **94**, 507–522 (1975).
- Di Chiara, G. & Imperato, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5274–5278 (1988).
- Jaber, M. *et al. Neuroscience* **65**, 1041–1050 (1995).
- Gerfen, C. R. *Trends Neurosci.* **15**, 133–138 (1992).
- Le Moine, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 230–234 (1990).
- Le Moine, C., Normand, E. & Bloch, B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4205–4209 (1991).
- Chiodo, L. A. *Neurochem. Int.* **26**, 815–845 (1992).
- Kawagoe, K. T., Zimmerman, J. B. & Wightman, R. M. *J. Neurosci. Meth.* **48**, 225–240 (1993).
- Wightman, R. M. & Zimmerman, J. B. *Brain Res. Rev.* **15**, 135–144 (1990).
- Jones, S. R., Garriss, P. A., Kilts, C. D. & Wightman, R. M. *J. Neurochem.* **64**, 2581–2589 (1995).
- Raiteri, M., Cerrito, F., Cervoni, A. M. & Levi, G. *J. Pharmac. exp. Ther.* **208**, 195–202 (1979).
- Wall, S. C., Gu, H. & Rudnick, G. *Molec. Pharmac.* **47**, 544–550 (1995).
- Parker, E. M. & Cubeddu, L. X. *J. Pharmac. exp. Ther.* **237**, 179–192 (1986).
- Jacocks, H. M. & Cox, B. M. *J. Pharmac. exp. Ther.* **262**, 356–364 (1992).
- Crow, T. J. *Br. J. Psychiat.* **137**, 383–386 (1980).
- Ehringer, H. & Hornykiewicz, O. *Klin. Wschr.* **38**, 1236–1239 (1960).
- Koob, G. F. & Bloom, F. E. *Science* **242**, 715–723 (1988).
- Graham, J. H., Maher, J. R. & Robinson, S. E. *J. Pharmac. exp. Ther.* **174**, 707–717 (1995).
- Tison, F., Normand, E. & Bloch, B. *Neurosci. Lett.* **166**, 48–50 (1994).
- Kennedy, R. T., Jones, S. R. & Wightman, R. M. *J. Neurochem.* **59**, 449–455 (1992).

ACKNOWLEDGEMENTS. We thank H. S. Kim for the mouse 129 SvJ genomic library; R. Hen for *neo*⁹ plasmid; B. Kohler for the production of embryonic stem cells and chimaeric mice; M. Hamon, M.-H. Thiebot, M.-P. Martres, C. Missale, P. Cox, E. H. Ellinwood and R. J. Lefkowitz for support and discussions; and S. Suter for technical help. B.G. received a NARSAD young researcher fellowship. M.J. received an EMBO long-term fellowship. S.J. received an NIA postdoctoral training fellowship. M.G.C. received a Bristol Myers Squibb unrestricted neuroscience award.

Galaxy harassment and the evolution of clusters of galaxies

Ben Moore*, Neal Katz*, George Lake*,
Alan Dressler† & Augustus Oemler Jr‡

* Department of Astronomy, University of Washington, Seattle, Washington 98195-1580, USA

† Observatories of the Carnegie Institution of Washington, Pasadena, California 91101-1292, USA

‡ Yale University Astronomy Department, New Haven, Connecticut 06511, USA

NEARBY clusters of galaxies are filled with red elliptical 'E' and lenticular 'S0' galaxies¹, while younger clusters (at redshifts of ≥ 0.4) contain substantial populations of blue spiral galaxies with morphological peculiarities²⁻⁷ (see Fig. 1). Thus, within the last 4–5 billion years, galaxies in clusters underwent strong evolution that completely changed their character. By contrast, galaxies that are not associated with clusters show far less morphological evolution⁸. Here we propose that multiple high-speed encounters between galaxies—'galaxy harassment'⁹—drives the morphological evolution in clusters. Our simulations show that these encounters are very different from mergers; they transform small disk galaxies into dwarf elliptical or dwarf spheroidal galaxies. Harassment will leave detectable debris arcs and could provide fuel for quasars in sub-luminous host galaxies.

Given a mechanism for distorting galaxies and promoting star formation that operates when a spiral first enters a cluster, hierarchical clustering models predict that most clusters will be filled with such galaxies at $z \approx 0.4$ (ref. 9). Proposed mechanisms include mergers^{10,11}, compression of gas in the high-pressure cluster environment^{12,13} and tidal compression by the cluster^{14,15}. Each of these scenarios can produce star-bursts, but none addresses morphological evolution or identifies the remnants of these distorted blue galaxies in present-day clusters. By analysing their Hubble Space Telescope (HST) images, Oemler *et al.*¹⁶

conclude that merging is implausible as the blue galaxy fraction is large and the merging probability is low. They observe disturbed spirals throughout the cluster, whereas both ram pressure stripping and global tides will operate efficiently only near the cluster's centre.

What mechanism drives star-bursts and rapid morphological evolution throughout a cluster of galaxies? Although direct mergers are extremely rare, every galaxy experiences a high-speed close encounter with a bright galaxy about once per Gyr. Here 'close' means within 50 kpc (several optical radii) and 'bright galaxy' is one as luminous as L_* , the characteristic break in the galaxy luminosity function¹⁷. (The Milky Way has a luminosity of $\sim L_*$.) To distinguish this from other collisional effects such as galaxy mergers and galaxy cannibalism, we refer to these frequent encounters as 'galaxy harassment'.

We use high-resolution numerical simulations to follow the evolution of a small bulgeless spiral orbiting within a dense cluster modelled on Coma (Fig. 1). Figure 2 shows the model galaxy along with an inventory of its stars, gas and dark matter. Some of our simulations use smoothed particle hydrodynamics¹⁸, to evolve the gas component of the disk at resolutions of 100–500 pc.

We simulated galaxies on circular and elliptical orbits in smooth cluster potentials before examining the effects of harassment. The disk galaxy shows little evolution over 5 Gyr when placed on a 450-kpc circular orbit in a smooth cluster potential. When placed on an eccentric orbit (apocentre of 300 kpc, pericentre at 600 kpc) the disk becomes bar unstable after the first pericentric passage¹⁴. The halo loses a small fraction of its mass at each pericentric passage, but stars and gas remain bound. The evolution is far more dramatic when we include the other cluster galaxies.

We have been conservative in our simulations to ensure that our results are robust. For a model galaxy on a fixed orbit, the havoc wreaked by harassment depends on the square of the masses of the largest galaxies encountered. While it has been speculated that the dark haloes are stripped within clusters¹⁹, we found that the dominant stripping mechanisms are tides and high-speed encounters: harassment²⁰. If galaxies are initially tidally limited by the cluster potential, bright galaxies retain more than half of the mass when followed for 5 Gyr.

The mean ratio of a perturbing galaxy's apocentre to its

FIG. 1 The dramatic evolution of clusters from $z \sim 0.4$ to today is shown in these images of their inner $300h^{-1}$ kpc, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. **a**, The nearby Coma cluster at $z = 0.023$ (courtesy of A. Fruchter, B. Moore and C. Steidel). Nearly every object surrounding the two central dominant elliptical galaxies is an E or S0. **b**, An HST image of CL0939 at $z = 0.42$ (ref. 5). This cluster is dominated by spiral galaxies; many appear disturbed yet no other galaxies are nearby. Although others seem normal, they have enhanced rates of star formation betrayed by strong $H\beta$ and $O II$ emission lines. The bright elliptical population is already in place at this redshift. The difference in galaxy populations is most dramatic at luminosities $> 2 \text{ mag}$ fainter than L_* . Fainter than this luminosity, 90% of galaxies in clusters at $z \approx 0.4$ are bulgeless 'Sd' disk systems, whereas 90% of galaxies within nearby clusters are dwarf ellipticals 'dE' or dE/S0 (refs 34, 35). The resolution of both images is $\sim 0.5h^{-1}$ kpc.

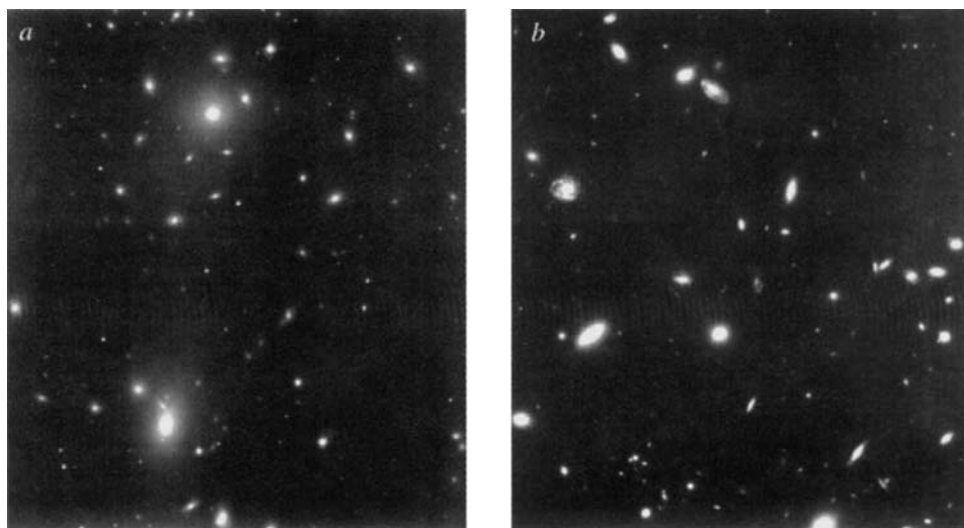
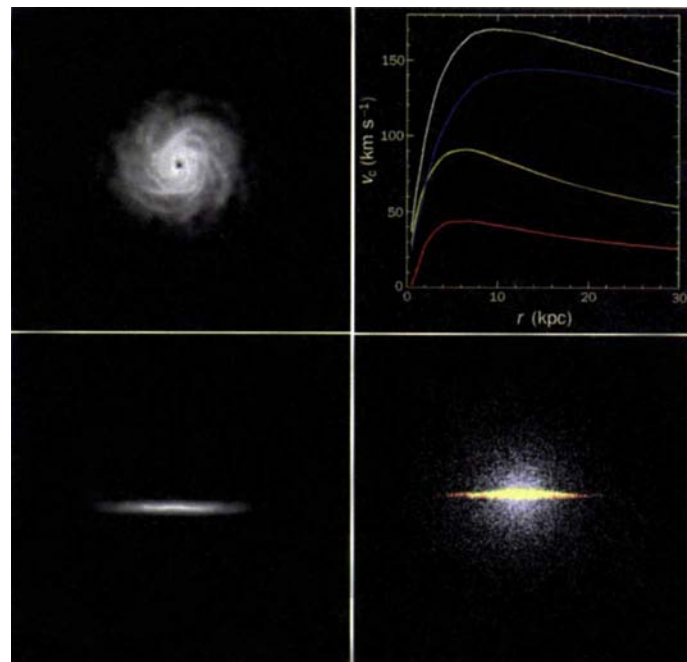


FIG. 2 The left panels show synthetic images of the initial state of our simulated galaxy viewed both edge and face on. In the simulated images, the stellar distribution is smoothed and filtered to model the resolution of the observations to a limiting brightness of 27 mag per square arcsecond. The bottom right panel shows the 3×2^{14} blue, red and yellow particles that make up the dark matter, gas and stars respectively. The exponential disk has a scale length of 2.5 kpc and a scale height of 200 pc. The disk is constructed with a Toomre 'stability' parameter $Q = 1.5$ and is run in isolation for 2 Gyr (as shown) before being set into orbit in the cluster. The gaseous disk is initially on cold circular orbits. The dark halo is a spherical isothermal with a core radius of 1 kpc that is tidally truncated at the pericentre of the galaxy's orbit. Within 20 kpc the ratio of dark matter to stars is 20:5:1 and the total mass is $\sim 10^{11} M_{\odot}$. The graph at the top right displays the contribution of each component to the rotational velocity of the disk. From the relationship found by Tully and Fisher³⁶, a circular velocity of $\sim 160 \text{ km s}^{-1}$ corresponds to a galaxy with a luminosity about $L/5$. We performed similar simulations with model galaxies with a peak circular velocity of $\sim 110 \text{ km s}^{-1}$. The galaxies are placed in a model of the Coma cluster that has a total mass within its virialized radius of $7 \times 10^{14} h^{-1} M_{\odot}$. Galaxies within the cluster are drawn from a Schechter luminosity function normalized to a typical cluster mass-to-light ratio of $250h$. This results in a cluster with 950 galaxies brighter than the Magellanic clouds, but only 31 brighter than L_{*} . Internal velocity dispersions are assigned by using the observed correlation with luminosity³⁷. The assigned masses assumed that the dark haloes are tidally truncated at their pericentric distance reduced by a time-averaged loss of 25% owing to harassment²⁰. The fraction of the cluster's density attached to galaxies varies from zero at its centre to nearly unity at its virial radius of $1.5h^{-1} \text{ Mpc}$. At $200h^{-1} \text{ kpc}$, roughly 15% of the mass is bound to galaxies. The rest of the cluster mass is in a smoothly distributed background represented by a fixed analytic potential. Further details can be found in Moore et al.²⁰.



pericentre in our cluster model is roughly 6:1; a galaxy with a mean radius of 400 kpc has a typical pericentre (r_{peri}) slightly greater than 150 kpc. We assign galaxies masses of $2.8 \times 10^{11} (r_{\text{peri}}/150 \text{ kpc})(L/L_{*})^{3/4} M_{\odot}$ corresponding to mass-to-light ratios, $M/L = 26h^2(r_{\text{peri}}/150 \text{ kpc})(L/L_{*})^{-1/4}$. The luminous parts of elliptical galaxies have mass-to-light ratios of $12hM_{\odot}/L_{*}$ (ref. 21), so our perturbing galaxies have modest amounts of dark matter.

At a fixed mean orbital radius, galaxies on elongated orbits experience greater harassment. We follow galaxies that have apocentre/pericentre ratios of 2 (that is, apocentre at 600 kpc, pericentre at 300 kpc). As a result these galaxies avoid extremes of the cluster distribution and start with large dark halo masses determined by the tidal limit at their atypically large pericentres. Both effects serve to underestimate the effects of harassment.

In Fig. 3, we compare the evolution of a harassed galaxy with images of galaxies in clusters. Typically, the first encounters create 'disturbed barred spirals' with sharp and dramatic features drawn out from the dynamically cold disk (Fig. 3a). Tails of material are distorted by the tidal field of the cluster (Fig. 3b). The gas distribution often forms ring structures that tumble within the stellar bar (Fig. 3c).

The evolution is driven by several close encounters that would stimulate the multiple starbursts inferred from HST data⁷. Another observational puzzle has been the ubiquity of disturbed galaxies with no sign of current interaction⁵. None of the images of our model galaxy has another cluster galaxy within 50 kpc. Over the course of 3 Gyr, the closest approach of another galaxy is $>30 \text{ kpc}$ away. Because the relative velocity of strong encounters is $\sim 1,500 \text{ km s}^{-1}$, and the velocity impulse internal to the galaxy is only $\lesssim 50 \text{ km s}^{-1}$, the perturbing galaxy moves $\sim 100 \text{ kpc}$ by the time that the disk's response is visible.

After several strong encounters, the loss of angular momentum to their own dark haloes and the perturbing galaxies, combined with impulsive heating, leads to a prolate figure flattened equally by random motions and rotation. The gas sinks to the centre of the galaxy and the stellar distribution is heated so that it closely

resembles a dwarf elliptical, although some retain very thick stellar disks and would be classed as dwarf lenticulars. At this stage in the evolution, encounters cease to create sharp distortions and fail to remove any more material from the compact remnant.

Using our simulations, we can identify the present-day remnants of the disturbed spirals seen at $z \approx 0.4$. Below L_{*} , two distinct classes of elliptical galaxy are observed. Low-luminosity Es with high central-surface brightness are a rare extension to the sequence of bright ellipticals; the archetype is M32. The most numerous galaxies in clusters are in a second class of dwarf ellipticals, also known as dwarf spheroidals (dE/dSph). Their exponential surface brightness profiles resemble those of spirals, as does the correlation of their low central-surface brightnesses with total luminosity. They are faint, at least 3 mag below L_{*} and as many as 14 mag if one extrapolates to the faintest known galaxies in the Local Group, Draco and Ursa Minor^{22,23}. Harassment transforms spirals into this latter class of galaxies. When allowing for mass loss (20–80%) plus 1–2 mag of fading from the cessation of star formation, our $L/5$ spiral galaxy becomes a dE/dSph galaxy that is 3–6 mag below L_{*} . The final one-dimensional velocity dispersions are in the range $30\text{--}60 \text{ km s}^{-1}$. The final stellar systems have surface density profiles, shapes and kinematics that are in good agreement with observations^{24–26}. Fig. 3d compares the remnant with a dwarf elliptical in the Coma cluster.

The observed stellar populations of dE galaxies implies recent star formation activity that can easily be understood in our model as a result of recent encounters with cluster galaxies²⁷. Harassed Sd spiral galaxies undergo a remarkable transformation from one morphological class to another without any merging taking place. Their dynamical states can account for all of the dissimilarities between dwarf elliptical and normal elliptical galaxies. Harassment provides the link between the dominant populations of galaxies in clusters at $z \approx 0.4$ and the present day.

Recent HST images of quasars suggest that many quasar hosts are not galaxies as luminous as L_{*} (ref. 28). This is surprising because simple energy considerations imply that quasars need $10^8\text{--}10^9 M_{\odot}$ to fuel the black-hole engine. To be conservative, one

prefers models where less than 10% of a galaxy's gas must be channelled to the centre. This argues for hosts as large as our own Milky Way, which has a few times $10^9 M_\odot$ of gas. From eight HST images of low-redshift quasars, Bahcall *et al.*²⁸ found that five of the host galaxies must be 0.5–1.5 mag fainter than L_* . Spiral galaxies of this luminosity have gas masses of order $10^8 M_\odot$; a quasar needs nearly all of it for fuel.

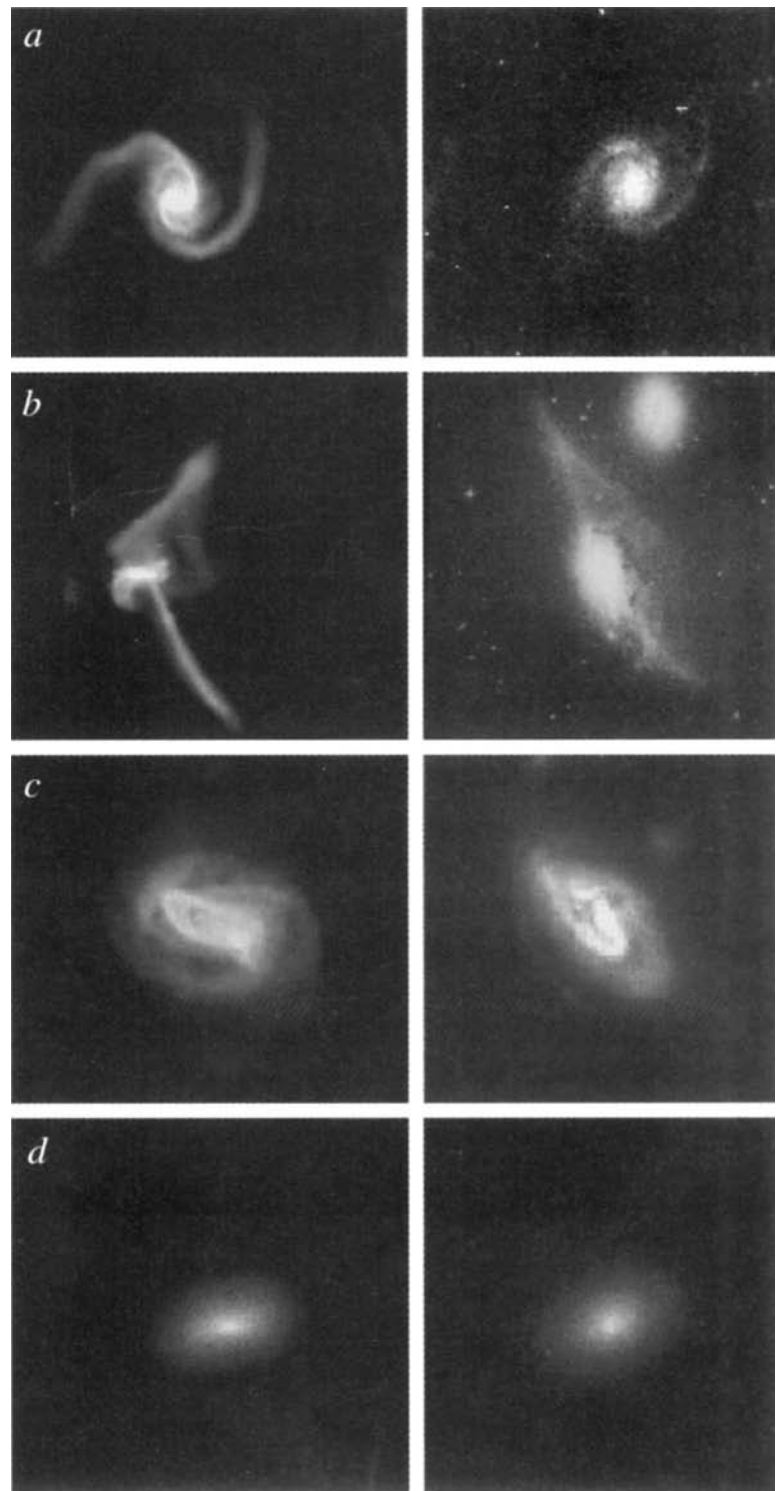
Galaxy harassment can give the quasar the fuel it needs. Although 10% of the gas is tidally stripped from the galaxy, the remaining 90% of the gas ($\geq 10^8 M_\odot$) sinks to the inner few hundred pc—the resolution of our simulation—by rapidly

losing angular momentum to the perturbing galaxies, dark halo and stellar bar.

Many quasars have luminous hosts²⁸. Also, quasars are known to avoid rich clusters²⁹, so how could they be harassed? Oddly, three of the five quasars with sub-luminous hosts found by Bahcall *et al.*³⁰ lie in clusters. A fourth is probably in a cluster and the environment of the fifth has not been studied. Harassed galaxies provide ideal hosts for quasars at intermediate redshifts known to lie in sub-luminous galaxies³¹.

Material torn from the disk creates debris tails that lead and follow the harassed galaxy. The tidal debris creates giant arcs that

FIG. 3 Comparisons of synthetic images from our simulations (left) with observations of harassed galaxies (right). *a*, The uppermost panels compare a disturbed spiral galaxy taken from the Butcher–Oemler cluster CL1447 with our model galaxy 150 Myr after suffering a single strong encounter that has pulled the two tails of material from the disk. At this time, the perturbing galaxy has already moved >200 kpc away. *b*, NGC4438, a spiral galaxy near the centre of the Virgo cluster (from Sandage and Bedke³⁸). The close companion in this image is probably not responsible for the disturbance³⁹. The model comparison is taken after 1 Gyr in the cluster. The tidal tails of material pulled from the galaxy have been subsequently tidally distorted. *c*, A spiral galaxy with a prominent ring in the distant rich cluster CL0939 is similar to the ring structures often observed in the harassment simulation. *d*, The bottom panels compare a dwarf elliptical galaxy in the Coma cluster with our model galaxy after 3 Gyr of evolution.



are thicker than the arcs caused by gravitational lensing. The debris arcs have redshifts that match the cluster and can have a wide range of shape owing to varying inclination. Harassment of the debris tails will create tidal shocks that promote the formation of dwarf galaxies, as seen in observations and simulations of tails in merging galaxies^{32,33}. Low surface-brightness features punctuated with dwarf galaxy formation should be detectable in HST images of $z \approx 0.4$ clusters. These slowly dissolve into a diffuse intra-cluster light that will provide another constraint on our evolutionary model. □

Received 23 June; accepted 29 December 1995.

1. Dressler, A. *Astrophys. J.* **236**, 351–365 (1980).
2. Butcher, H. & Oemler, A. *Astrophys. J.* **219**, 18–33 (1978).
3. Butcher, H. & Oemler, A. *Astrophys. J.* **285**, 426–438 (1984).
4. Dressler, A., Oemler, A., Butcher, H. & Gunn, J. E. *Astrophys. J.* **430**, 107–120 (1994).
5. Dressler, A., Oemler, A., Sparks, W. B. & Lucas, R. A. *Astrophys. J.* **435**, L23–L26 (1994).
6. Couch, W. J., Ellis, R. S., Sharples, R. & Smail, I. *Astrophys. J.* **430**, 121–138 (1994).
7. Barger, A. J. *et al.* *Mon. Not. R. astr. Soc.* (in the press).
8. Griffiths, R. E. *et al.* *Astrophys. J.* **435**, L19–L22 (1994).
9. Kauffmann, G. *Mon. Not. R. astr. Soc.* **274**, 153–160 (1995).
10. Icke, V. *Astr. Astrophys.* **144**, 115–123 (1985).
11. Miller, R. H. *Comments Astrophys.* **13**, 1–11 (1988).
12. Dressler, A. & Gunn, J. E. *Astrophys. J.* **270**, 7–19 (1983).

13. Evrard, A. E. *Mon. Not. R. astr. Soc.* **248**, 8p–10 (1991).
14. Byrd, G. & Valtonen, M. *Astrophys. J.* **350**, 89–94 (1990).
15. Valluri, M. *Astrophys. J.* **408**, 57–70 (1993).
16. Oemler, A., Dressler, A. & Butcher, H. R. *Astrophys. J.* (in the press).
17. Schechter, P. *Astrophys. J.* **302**, 297–306 (1976).
18. Hernquist, L. & Katz, N. *Astrophys. J. Suppl.* **70**, 419–446 (1989).
19. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341–358 (1978).
20. Moore, B., Katz, N. & Lake, G. *Astrophys. J.* **457**, 455–458 (1996).
21. van der Marel, R. P. *Mon. Not. R. astr. Soc.* **253**, 710–726 (1991).
22. Kormendy, J. & Bender, R. in *ESO/DHP Symposium on Dwarf Galaxies* (eds Meylan, G. & Prugniel, P.) (European Southern Observatory, Munich, in the press).
23. Ferguson, H. C. & Binggeli, B. *Astr. Astrophys. Rev.* **6**, 67–122 (1994).
24. Bender, R. & Nieto, J. L. *Astr. Astrophys.* **239**, 97–112 (1990).
25. Faber, S. M. & Lin, D. N. C. *Astrophys. J.* **266**, L17–L21 (1983).
26. Binggeli, B. & Popescu, C. G. *Astr. Astrophys.* **298**, 63–76 (1994).
27. Held, E. V. & Mould, J. R. *Astr. J.* **107**, 1307–1319 (1994).
28. Disney, M. J. *et al.* *Nature* **376**, 150–153 (1995).
29. French, H. B. & Gunn, J. E. *Astrophys. J.* **269**, 29–34 (1983).
30. Bahcall, J. N., Kirkhakis, S. & Schneider, D. P. *Astrophys. J.* **435**, L11–L14 (1994).
31. Lake, G., Katz, N. & Moore, B. *Astrophys. J.* (in the press).
32. Barnes, J. E. & Hernquist, L. *Nature* **360**, 715–717 (1992).
33. Mirabel, I. F., Dottori, H. & Lutz, D. *Astr. Astrophys.* **256**, L19–L22 (1992).
34. Binggeli, B., Tammann, G. A. & Sandage, A. *Astr. J.* **94**, 251–277 (1987).
35. Thompson, L. A. & Gregory, S. A. *Astr. J.* **106**, 2197–2212 (1993).
36. Tully, R. B. & Fisher, J. R. *Astr. Astrophys.* **54**, 661–674 (1977).
37. Faber, S. M. & Jackson, R. E. *Astrophys. J.* **204**, 668–683 (1976).
38. Sandage, A. & Bedke, J. *The Carnegie Atlas of Galaxies* (Carnegie Inst. Washington, 1994).
39. Combes, F., Dupraz, C., Casoli, F. & Pagani, L. *Astr. Astrophys.* **203**, L9–L12 (1988).

ACKNOWLEDGEMENTS. This research was funded by NASA through the LTSA, ATP and HPCC/ESS programmes.

Microscopic origins of superfluidity in confined geometries

P. E. Sokol*†, M. R. Gibbs†, W. G. Stirling†, R. T. Azuah‡ & M. A. Adams‡

* Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

† Department of Physics, Keele University, Keele, Staffordshire ST5 5BG, UK

‡ ISIS Division, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

LIQUID helium provides a convenient model system in which to study the effects of disorder on strongly interacting media such as superfluids^{1,2} and superconductors³. Confinement of liquid helium in porous glasses profoundly affects its behaviour, even to the extent of changing the universality class of the normal-to-superfluid phase transition^{1,4,5}. Although the effects of disorder on macroscopic fluid properties (such as the superfluid fraction) have been studied extensively^{6–11}, the microscopic processes underlying these effects have received much less attention. For example, little is known theoretically¹² or experimentally^{13,14} about the effects of disorder on the spectrum of elementary superfluid excitations, which reflects the dynamics of the system on a microscopic scale. Here we report inelastic neutron scattering measurements of the collective excitation spectrum for ⁴He confined in porous aerogel glass. Near the superfluid transition temperature, the behaviour of the superfluid phase is governed by rotons (elementary excitations that are often compared to microscopic vortex rings), which we find to exhibit an increased effective mass and a decreased lifetime in the disordered system, relative to the unconfined superfluid. No theoretical predictions for these changes exist, and their origin is unclear; nevertheless, the disorder-induced changes in the microscopic collective excitation spectrum account fully for the observed changes in macroscopic fluid behaviour.

The exact nature of the confining medium determines the type of disorder and its effects on the transition. For example, the superfluid transition of helium confined in vycor glass, which has a compact structure with well defined small (~ 70 -Å) pores, is

considerably suppressed, but still has the same critical exponent (0.674) indicating that it remains in the same universality class^{1,2}. Alternately, confinement in aerogel (Fig. 1), a tenuous porous glass with a broad distribution of pore sizes, barely changes the transition temperature while drastically changing the critical exponent (0.813) indicating that the system is in a new universality class^{1,5}. Additionally, at lower temperatures, outside the critical region, the superfluid fraction¹⁵ is enhanced.

Here we report neutron inelastic scattering measurements of the excitation spectrum of superfluid helium confined in aerogel¹⁶ which explore the microscopic effects of this disorder¹⁷. The measurements were carried out at the Rutherford Appleton Laboratory using the IRIS spectrometer at the ISIS pulsed neutron source. With a fixed final energy of 5 meV, this spectrometer covers momentum transfers of $0.3 < Q < 2.1 \text{ Å}^{-1}$, spanning the phonon–maxon–roton region of the ⁴He dispersion curve. The (elastic) energy resolution was 15 μ eV. The aerogel sample (91% porosity) was filled to 95% of full pore capacity with high-purity ⁴He gas. Measurements were performed at five temperatures between 1.3 and 2.3 K, as determined by a germanium resistance thermometer, with a temperature stability of ± 0.015 K. The neutron data were converted to the dynamic scattering function S , which is a function of momentum (Q) and energy (E) transfer, using standard techniques^{18,19}. The dynamic scattering function is directly proportional to the space and time Fourier transform of the density–density correlation function and provides direct information on the collective excitations that determine the properties of the liquid.

The dynamic scattering function of the ⁴He confined in aerogel was quite similar to that observed for bulk ⁴He. S exhibited the familiar phonon–maxon–roton dispersion relation for the elementary excitations which are characteristic of ⁴He in the superfluid phase. This excitation curve is usually separated into two components. The characteristic excitations at low momentum are known as phonons, which represent long-wavelength density fluctuations (like sound waves in a solid). The characteristic excitations at intermediate momenta are known as rotons, which have historically been viewed as microscopic vortex rings with a core just large enough for a single atom to pass through. More modern theories treat the roton as a low-energy short-wavelength density excitation, characteristic of the superfluid phase, with a wavelength characteristic of the interatomic spacing of the liquid. The single phonon excitations were well defined and could be least-squares fitted to a single damped harmonic oscillator expression²⁰, which provides an excellent description of the bulk ⁴He

data, to obtain their energy and width (inverse lifetime). The excitation energies in the vicinity of the roton, which dominates the superfluid thermodynamics above 1 K, are given by the Landau expression²¹

$$E(Q) = \Delta + \frac{\hbar^2(Q - Q_R)^2}{2\mu}$$

where $\hbar Q_R$ is the roton momentum, μ is the roton effective mass and Δ is the roton energy gap. The excitation energies for ^4He in aerogel at several different temperatures are presented (data points) as a function of Q in Fig. 2. As shown in the figure, this form provides an excellent description of the experimental results; the parameters used to fit the data are given in Table 1. The values for the bulk liquid are also shown in the figure (dotted line, 1.38 K; dashed line, 1.9 K) for comparison²².

The differences in the microscopic dynamics of the bulk and confined liquids are clearly apparent. At low temperature (1.3 K) the roton energy gap for the two liquids is the same, while the roton effective mass of the confined liquid is slightly larger. However, at higher temperatures, the bulk liquid exhibits a much larger temperature dependence, at least up to 1.9 K, than the helium in aerogel. It is these differences give rise to the different superfluid behaviour in the bulk and confined liquids.

The well known two-fluid model²³ of superfluid ^4He describes the total fluid density as the sum of a superfluid density ρ_s and a normal fluid density ρ_n which together comprise the total mass of the liquid, ρ . The normal fluid component, in the superfluid, arises from elementary excitations (phonons and rotons) exchanging momentum with the walls of the container. Thus, the normal fluid density can be directly related to the spectrum of elementary excitations and will be determined by the number of excitations that are thermally activated. At temperatures above 1 K, the phonon excitation makes a negligible contribution; ρ_s/ρ is dominated by the roton and is given by²³

$$\frac{\rho_s}{\rho} = 1 - \frac{2\hbar\mu^{1/2}Q_R^4}{\rho^3(2\pi)^{3/2}(k_B T)^{1/2}} e^{-\Delta/k_B T}$$

In the bulk liquid this expression provides an excellent description²² of the macroscopically determined²⁴ ρ_s/ρ , as shown in Fig. 1.

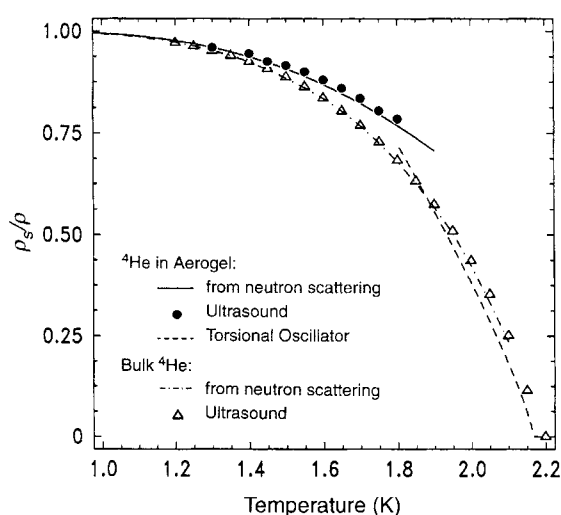


FIG. 1 The superfluid fraction of liquid helium in the bulk and confined in aerogel. For liquid helium confined in aerogel the measurements are: solid line, predictions based on the measured microscopic excitations spectrum; filled circles, ultrasonic measurements¹⁵; dashed line, torsional oscillator measurements near the transition temperature¹. The bulk measurements are: dash-dot line, predictions based on the microscopic excitations²²; triangles, ultrasonic measurements²⁴.

TABLE 1 Parameters used to fit data in Fig. 2

Temp. (K)	Q_R ($\text{\AA}^{-1}$)	Δ (meV)	μ_R (^4He mass)
1.3	1.926 ± 0.002	0.742 ± 0.001	0.167 ± 0.012
1.4	1.923 ± 0.003	0.738 ± 0.002	0.159 ± 0.014
1.7	1.927 ± 0.002	0.731 ± 0.001	0.155 ± 0.011
1.9	1.929 ± 0.002	0.726 ± 0.002	0.148 ± 0.010

The roton parameters for ^4He in aerogel may also be used to calculate ρ_s/ρ , as in the bulk. Figure 1 shows a comparison of the measured values, via ultrasonic¹⁵ and torsional oscillator¹ measurements, and those calculated from the microscopic excitation spectrum. The coarse-grained density of the liquid in the pores, defined as the number of helium atoms divided by the total sample volume, has been used. The results correctly mirror the enhancement of ρ_s/ρ below 1.9 K and are in excellent quantitative agreement. Furthermore, ρ_s/ρ at the superfluid transition temperature, T_λ , is determined only by the roton energy gap. If ρ_s/ρ is to go to zero, as the two-fluid model requires, then the energy gap for the helium in aerogel must be comparable to that of the bulk liquid, as their densities and transition temperatures are nearly the same. Thus, Δ must exhibit a much stronger temperature dependence in the critical region, giving rise to the different critical exponents¹. In general, it is the dimensionality of the system, the degrees of freedom of the order parameter, and the nature of the disorder that determine the universality class and critical behaviour for the system. Our results show that these macroscopic dynamics can be directly related to the microscopic excitations of the system in the presence of disorder.

The above results demonstrate the microscopic origin of the macroscopic changes induced by confinement. However, the origins of the changes in the microscopic dynamics are still virtually unexplored. Theoretical studies¹² have predicted a change in the low-energy excitations. However, no theoretical understanding of the effects of disorder on the roton, which dominates the superfluid behaviour, is available at present. An examination of the widths of the excitations, shown in Fig. 3,

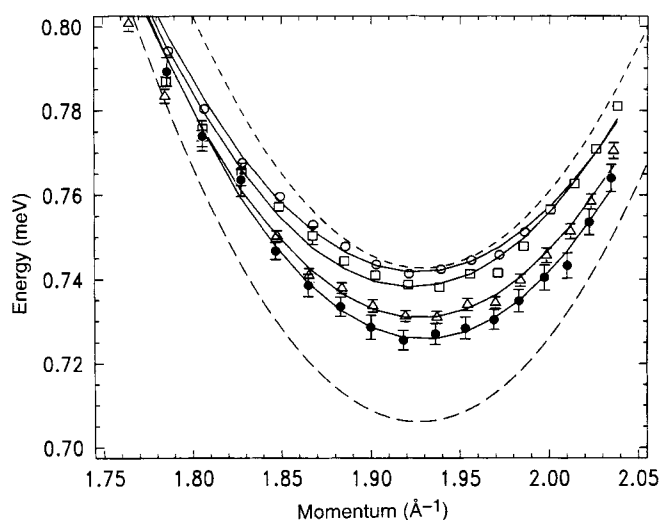


FIG. 2 Excitation energies of ^4He in aerogel in the roton region at: open circles, 1.3 K; open squares, 1.4 K; open triangles, 1.7 K; and filled circles, 1.9 K. The solid lines show the best least-squares fit to the experimental data using the Landau expression; the values of the parameters in each fit are shown in Table 1. The excitation energies in the roton region for the bulk liquid at 1.38 and 1.9 K are also shown²² as dotted and dashed lines respectively.

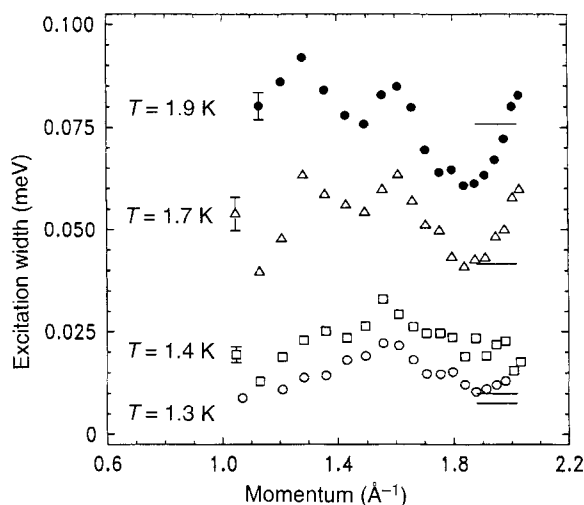


FIG. 3 Momentum transfer dependence of the excitation width (half width at half maximum HWHM) at several temperatures below T_λ . Open circles, 1.3 K; open squares, 1.4 K; open triangles, 1.7 K; and filled circles, 1.9 K. The solid lines are the corresponding values at the roton minimum ($Q = 1.92 \text{ \AA}^{-1}$) in the bulk²⁵.

illustrates this point more clearly. The excitations in the confined liquid are considerably broader than those in the bulk²⁵ (that is, they have a shorter lifetime) and exhibit a different momentum and temperature dependence. No theoretical predictions for these changes exist and their origin is still unclear.

The broadening of the roton is unlikely to arise from finite-size effects. The relatively large pores ($\sim 500 \text{ \AA}$) in aerogel would yield a broadening that would be unobservable in our measurement (the finite size broadening is of the order of $0.7 \text{ \mu eV}$, compared to our instrumental resolution of $15 \text{ \mu eV}$). An alternative broadening mechanism may arise from trapped vorticity in the disordered geometry. Recent theories of the superfluid transition indicate that the appearance of thermally activated vortex lines play a major role in the superfluid transition. In addition, it is known that vortex pinning is also important in high- T_c superconductors. Confinement of superfluid helium leads to remnant vortex lines even when no flow is present²⁶. If a sufficiently high density of vortex lines were trapped roton-vortex scattering events could lead to a decrease in the roton lifetime²⁷ and an increase in the roton effective mass, owing to the difficulty of moving through the vortex tangle. The effect of trapped vorticity on the temperature dependence of the roton gap, which governs the behaviour of the superfluid near the transition, is not clear. But strong roton-vortex scattering, which may depend little on temperature, may mask other effects and lead to the smaller observed temperature dependence. Further studies, both experimental and theoretical, into the effects of disorder and the role of vortices in the superfluid and superconducting transitions will be required. $\square$

Received 20 July 1995; accepted 18 January 1996.

- Chan, M. H. W., Blum, K. I., Murphy, S. Q., Wong, G. K. S. & Reppy, J. D. *Phys. Rev. Lett.* **61**, 1950–1953 (1988).
- Berthold, J. E., Bishop, D. J. & Reppy, J. D. *Phys. Rev. Lett.* **39**, 348–351 (1977).
- Tyler, A., Cho, H. A. & Reppy, J. D. *Phys. Rev. Lett.* **73**, 987–990 (1994).
- Wong, G. K. S., Crowell, P. A., Cho, H. A. & Reppy, J. D. *Phys. Rev. Lett.* **65**, 2410–2413 (1990).
- Wong, G. K. S., Crowell, P. A., Cho, H. A. & Reppy, J. D. *Phys. Rev. Lett.* **68**, 3858–3860 (1992).
- Crowell, P. A. et al. *Phys. Rev.* **B51**, 12721–12736 (1995).
- Harris, A. B. *J. Phys.* **C7**, 1671–1674 (1974).
- Weinreb, A. & Halperin, B. I. *Phys. Rev.* **B27**, 413–427 (1983).
- Narayan, O. & Fisher, D. S. *Phys. Rev.* **B42**, 7869–7875 (1990).
- Hohenberg, P. C., Aharony, A., Halperin, B. I. & Siggia, E. D. *Phys. Rev.* **B13**, 2986–2996 (1976).
- Stauffer, D., Ferer, M. & Wortis, M. *Phys. Rev. Lett.* **29**, 345–348 (1972).
- Makivic, M., Trivedi, N. & Ullah, S. *Phys. Rev. Lett.* **71**, 2307–2310 (1993).
- De Kinder, J., Coddens, G. & Millet, R. *Z. Phys.* **B95**, 511–514 (1994).
- Snow, W. M. & Sokol, P. E. *J. Low Temp. Phys.* **80**, 197–220 (1990).

- McKenna, M. J., Slawicki, T. & Maynard, J. D. *Phys. Rev. Lett.* **66**, 1878–1881 (1991).
- Aerogels: Proceedings of the First International Symposium* (Springer-Verlag, Berlin, 1986).
- Gibbs, M. R., Sokol, P. E., Azuah, R. T., Stirling, W. G. & Adams, M. A. *Physica* **B213&214**, 462–464 (1995).
- Price, D. L. & Skold, K. in *Methods of Experimental Physics* (eds Skold, K. & Price, D. L.) **1** (Academic, New York, 1986).
- Osborn, R., Ecclestone, R. S., Benington, S. M. & Kitchener, B. G. B. *Report RAL-94-102* (Rutherford Appleton Lab., Chilton, 1995).
- Stirling, W. G. & Glyde, H. R. *Phys. Rev.* **B41**, 4224–4242 (1990).
- Wilks, J. & Betts, D. S. *An Introduction to Liquid Helium* (Clarendon, Oxford, 1987).
- Brooks, J. S. & Donnelly, R. J. *J. Phys. Chem. Ref. Data* **6**, 51–104 (1977).
- Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* (Institute of Physics Publishing, Bristol, 1990).
- Maynard, J. D. *Phys. Rev.* **B14**, 3868–3891 (1976).
- Andersen, K. H. & Stirling, W. G. *J. Phys.: Condens. Matter* **6**, 5805–5822 (1994).
- Awschalom, D. D. & Schwarz, K. W. *Phys. Rev. Lett.* **52**, 49–52 (1984).
- Reatto, L., Vitiello, S. A. & Masserini, G. L. *J. Low Temp. Phys.* **93**, 879–892 (1993).

ACKNOWLEDGEMENTS. We thank the staff of the Rutherford Appleton Laboratory for their assistance. This work was supported by the US National Science Foundation and the UK Engineering Science Physical Research Council.

Characterization of low-dimensional dynamics in the crayfish caudal photoreceptor

Xing Pei & Frank Moss

Laboratory for Neurodynamics, Department of Physics and Astronomy, University of Missouri at St Louis, St Louis, Missouri 63121, USA

ATTEMPTS to detect and characterize chaos in biological systems are of considerable interest, especially in medical science, where successful demonstrations may lead to new diagnostic tools and therapies¹. Unfortunately, conventional methods for identifying chaos often yield equivocal results when applied to biological data^{2–8}, which are usually heavily contaminated with noise. For such applications, a new technique¹ based on the detection of unstable periodic orbits holds promise. Infinite sets of unstable periodic orbits underlie chaos in dissipative systems^{4,9}; accordingly, the new method searches a time series only for rare events⁸ characteristic of these unstable orbits¹⁰, rather than analysing the structure of the series as a whole. Here we demonstrate the efficacy of the method when applied to the dynamics of the crayfish caudal photoreceptor (subject to stimuli representative of the animal's natural habitat). Our findings confirm the existence of low-dimensional dynamics in the system, and strongly suggest the existence of deterministic chaos. More importantly, these results demonstrate the power of methods based on the detection of unstable periodic orbits for identifying low-dimensional dynamics—and, in particular, chaos—in biological systems.

Since the earliest demonstrations of chaos in physical nonlinear systems of three or more dimensions^{11–13}, the question of its existence in biological systems, which are certainly nonlinear and multidimensional, naturally arose^{14,15}. But in such cases one never has an accurate model for the dynamics. It is therefore necessary to devise reliable tests for chaos using only the output record of such systems. Much effort has been put into this enterprise¹⁶, mostly inspired by original tests, useful for most numerical and many physical systems^{17,18}. However, in order to give accurate and reliable results, these methods, as well as their many variants, must be applied to data files which are long and relatively free of noise contamination. Unfortunately, living systems are both noisy and changeable with time, so that only relatively short files can be obtained under constant conditions. Thus the detection of low-dimensional deterministic behaviour or chaos (the distinction is discussed below) in biological systems using traditional methods has frequently been questioned^{2,3}. Some examples, however, have provided evidence for deterministic

behaviour in biological data using both linear and nonlinear forecasting techniques^{6,7,19}.

Recently, a new technique has been used for control of chaos in both heart and brain preparations^{1,20}. In these studies, however, the emphasis was on control rather than on the efficacy of the technique at detecting low-dimensional behaviour, and the effects of random processes were not addressed. But such processes must be important, as some degree of control of the dynamics can be achieved in even randomly forced non-chaotic systems using the same algorithms²¹. The questions raised by these observations can only be answered by statistical studies.

Here we show statistically significant evidence of the existence of unstable periodic orbits (UPOs) in the caudal photoreceptor (CPR) neuron in the sixth ganglion of the crayfish *Procambarus clarkii*^{22,23}. The hydrodynamically sensitive hair mechanoreceptors on the tailfan were periodically stimulated by small-amplitude relative fluid motions²⁴, while the CPR was exposed to a steady light. UPOs were observed for certain combinations of three parameters: stimulus amplitude, stimulus frequency and light intensity, in ranges typically found in the animal's natural environment. This stands in contrast to previous experiments wherein the

putative chaos was artificially induced either chemically^{1,20}, or with intracellular electrical stimuli²⁵.

The method rests on the detection of "exceptional events"⁸, that is, events which may occur rarely in a data file, but nevertheless signal the existence of deterministic behaviour with some statistical reliability. In the present case, an "event" is an encounter with a UPO, as defined by a criterion outlined below. Though the UPO may be visited rarely in a given file, once the trajectory falls on it, a predictable pattern of time intervals follows. The statistical significance of the occurrence frequency of this pattern can be tested by comparison with a random, or surrogate, file. Files of time intervals between recorded action potentials are searched for such events, which are extracted for analysis, while the remainder of the file is ignored. By contrast, traditional methods analyse the complete file including the noise, and this, perhaps, explains why the present method can detect UPOs in quite noisy records¹⁰. Because the records are both noisy and of relatively short length, the reliability of detection becomes a matter of statistical measure. We count the number of events within a data file and compare it to the number found in a surrogate, which is a randomized representation of the original data file⁵. A statistical measure is then given by

$$K = \frac{N - \langle N_s \rangle}{\sigma} \quad (1)$$

where N is the number of events found in the data file, $\langle N_s \rangle$ is the mean number of events detected in different realizations of the surrogate files, and σ is the standard deviation of the N_s . Thus K measures the distance between the real finding and the mean surrogate finding in units of the standard deviation of the surrogate findings. Assuming gaussian statistics, $K \geq 2$ represents a determination with a confidence level greater than 95% (ref. 26). In a noisy file, the actual values of N and N_s depend on the stringency of the criterion used to define an event¹⁰, an optimized version of which might lead to improved detection statistics in future applications. Optimization may also include different methods for generating surrogates⁵.

That chaos can be detected, and hence controlled, by searching for signatures of UPOs has been previously demonstrated in a variety of physical²⁷⁻²⁹ and chemical^{30,31} systems. It is based on the theory that dissipative chaos is built upon the skeleton of a countable infinity of such orbits^{32,33}, a view which is also central to chaos control theory³⁴. The statistics of this method and its robustness to noise have been previously studied using files generated with a noise-driven, chaotic electronic circuit designed to mimic typical biological data¹⁰.

An encounter, or event, is defined by the following criterion: (1) three points approaching the line of identity with sequentially decreasing perpendicular distances followed by three departing with sequentially increasing distances; (2) straight lines matched to the two sets (linear least-squares) must have slopes m in the ranges $0 \geq m_s > -1$ and $-1 \geq m_{us} > -\infty$ (subscripts identify stable and unstable directions) and (3) the intersection must lie within a distance to the line of identity smaller than half the mean of the five distances in the two sets (Fig. 1). There are only five distances, because one point (the closest to the period-one point) is common to both sets. The slope conditions apply to "flip" saddles, that is, those for which the sequence of points alternate between the sides of the line of identity, Fig. 1a. Nonflip saddles are dynamically possible but have not yet been found in biological preparations¹.

From each recording of action potentials, the interspike time intervals are assembled into a scatter plot (Fig. 1b). This record is then searched for the number of encounters, N , which satisfy the aforementioned criterion. Thereafter, 100 surrogate records are constructed using 100 different realizations of the randomization process. Each realization is searched for the number of encounters, N_s , satisfying the same criterion. The statistic, K , is then obtained according to equation (1). An example set of selected encounters of original data and one of their surrogates are shown in Fig. 1c

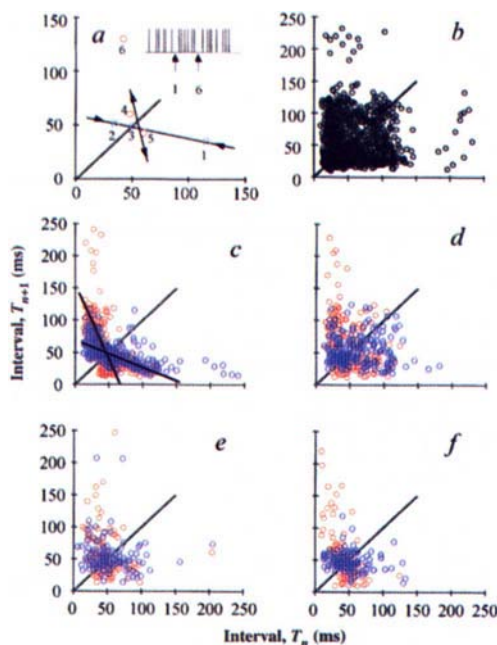


FIG. 1 Time-interval scatter plots from extracellular recordings on the caudal photoreceptor of crayfish *P. clarkii*. A time interval between action potentials T_{n+1} , is plotted against the preceding interval T_n . a, An example of an unstable periodic orbit, (UPO) observed as a numbered (1–6) sequence of points (blue) approaching the line of identity (45° line) with ever-decreasing separations followed by a sequence of points (red) departing at ever-increasing distances¹. The arrows identify this sequence on the physiological recording (inset). The magnitudes of the eigenvalues of the linearized dynamics near a UPO are given by the inverse slopes of straight lines matched (using linear least-squares) to the approaching and departing sequences. The intersection locates the neighbourhood of the period one fixed point. This definition only identifies UPOs with one real positive and one real negative eigenvalue. It discriminates against the fixed points arising from stable limit cycles, for which at low noise $K < 0$ (ref. 10). We consider here only two-dimensional projections and period-one orbits. b, An example of a scatter plot, for parameters as follows: stimulus amplitude, 40 dB; stimulus frequency, 8.5 Hz, light level, 0.15 $\mu\text{W mm}^{-2}$. The complete file had 7,860 intervals of which only the first 1,000 are shown here. c, Selected encounters ($N = 138$) from the complete file showing stable (blue points) and unstable (red points) mean eigendirections (straight lines). d, Encounters ($N_s = 76$) selected from one surrogate of b. The file shown in b, c and d represents an example for large measure, $K = 9.4$. e and f show a small-measure example, $K = 0.5$, obtained for smaller stimulus amplitude (48 dB).

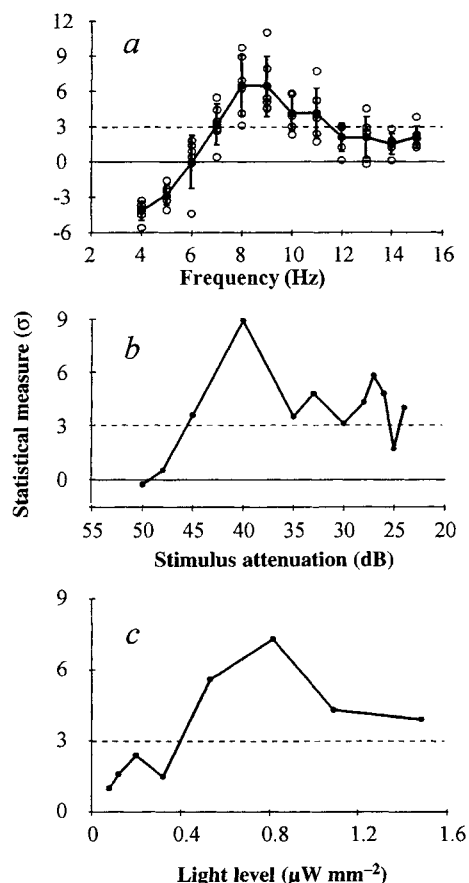


FIG. 2 The statistical measure K versus the three parameters employed in the experiments. The dashed lines mark $K = 3$. The data are from three cells of a total of 24 tested, all of which showed evidence of UPOs for some parameter values, with the strongest influence being the light intensity. The intact tailfan and sixth ganglion were excised under standard Crayfish Ringer solution and mounted on a transducer which provided periodic relative motion between the hairs and the bath fluid in electrical and vibration isolation (Technical Manufacturing Corp.). A standard attenuated sinusoidal voltage (Hewlett Packard 355) drove the transducer. Calibrated (Graseby 371), steady light (tungsten spectrum, CHIU Tech Lumina) was introduced into the preparation by fibre optics. Neural spikes were selected (WPI 121), sampled at 10 kHz (CED 1401 plus) and analysed (CED Spike 2). Typical recordings varied from 3,000 to 8,000 time intervals of which about 0.8–3%, depending on the K value, comprised selected encounters. Values $K \geq 3$ (dashed lines) identify the presence of a UPO ($\sim 99\%$ confidence). *a*, K versus the stimulus frequency. Six sets of measurements (open circles, same cell; amplitude, 30 dB; light level, $0.2 \mu\text{W mm}^{-2}$) show the mean K (solid line), and standard deviations (error bars) versus frequency. *b*, K versus stimulus amplitude (frequency, 8.5 Hz; light level, $0.12 \mu\text{W mm}^{-2}$). Both displacement and velocity of the preparation relative to the solution were calibrated with a laser Doppler vibrometer (Polytec OFV 501) versus signal attenuation (dB) for a standard driving voltage. Two examples of calibration points at 10 Hz are: 35 dB, $2.39 \mu\text{m}$ peak displacement, $158 \mu\text{s}^{-1}$ peak velocity; and 55 dB, $0.255 \mu\text{m}$, $15.0 \mu\text{s}^{-1}$. *c*, K versus light level (frequency, 8.0 Hz; amplitude, 40 dB). Our observations failed to detect determinism in any cell tested in the dark (light level $< 0.001 \mu\text{W mm}^{-2}$) or in light which was too intense ($> 2 \mu\text{W mm}^{-2}$).

and *d* respectively, for large K . The points in Fig. 1c clearly form two distinct groups with well defined eigendirections (straight lines with negative slopes). The morphology of the surrogate encounter points (Fig. 1d) is different from that found in the original data, even though the encounter selection criterion was identical: no statistically significant eigendirections are evident. The statistical measure, K , quantifies this difference. For comparison, a small- K example is shown in Fig. 1e and f, where original

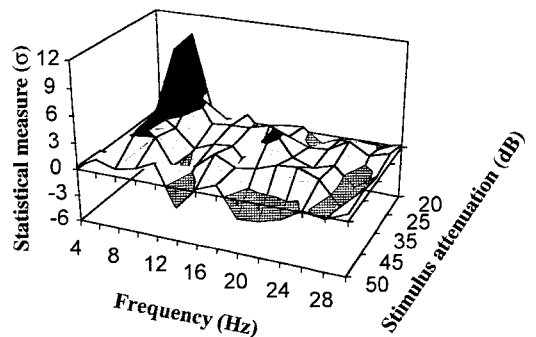


FIG. 3 Data from a single cell displayed in three dimensions for constant light level ($0.1 \mu\text{W mm}^{-2}$). Regions with $K \geq 3$ are shown by the darkest shade. The detailed shape of this surface changes with the individual cell. No deterministic behaviour was observed for parameter values much outside the ranges shown. Different CPR cells (different animals) resulted in different shaped curves, but all cells of the 24 tested showed broad maxima with $K > 3$ lying within ranges of the parameters similar to those shown. The major features of this surface are governed by the light level, with higher light levels resulting in greater complexity (more peaks above $K = 3$). The example shown is for relatively low light level and shows only two peaks (dark shades). The hatched regions identify $K < 0$, which, in our previous work¹⁰, indicated the presence of limit cycles.

and surrogate data show similar morphologies, and no eigen-directions are evident.

Figure 2 shows the behaviour of K as the three parameters are varied in turn; stimulus frequency (at constant amplitude and light intensity; Fig. 2a); amplitude (at constant frequency and light intensity; Fig. 2b); and light intensity (at constant frequency and amplitude; Fig. 2c). Representative data from another cell are displayed in Fig. 3. Figures 2 and 3 both show that the system enters into and departs from regions exhibiting UPOs as each of the parameters in turn is varied over its range, a behaviour typical of forced nonlinear oscillators^{35,36}.

These data show that the lowest-order (period-one) UPOs can be detected in noisy biological files from the periodically forced crayfish CPR neuron with large statistical confidence levels. We have not detected chaos with certainty in this experiment, though the eigenvalues and saddle type (flip) found here are comparable to those found in a known chaotic system contaminated with noise¹⁰. Theorists point out that the detection of one member of an infinite set does not constitute an existence proof of the set, and that UPOs can exist in non-chaotic systems. We have detected UPOs of order two and three (data not shown here) and made higher-dimensional embeddings, but the statistical confidence levels of those determinations degrade rapidly for higher orders. UPOs of the lowest orders seem to be well separated but become more dense as the order increases. Thus the noise blurs higher-order separations so that, in this experiment at least, it was not possible to observe orders higher than three. This dims the prospect that a scaling law for the infinite set of UPOs, which would address the aforementioned theoretical concerns, might be experimentally obtained.

Why should the crayfish CPR exhibit UPOs? We believe it to be an example of a periodically forced nonlinear oscillator^{35,36}. It is well known that the CPR exhibits rhythmic pacemaker activity which becomes more evident when all sensory input from the hair mechanoreceptors is cut³⁷. For appropriate parameter values, periodic forcing of the hair receptors may drive this oscillator into chaos, giving rise to the UPOs observed here. The function of the light level is less clear, though it has been shown to mediate signal processing in the CPR cell (X.P., L. Wilkens and F.M., unpublished data). It is possible that the animal makes use of the observed low-dimensional deterministic dynamics reported here,

because the parameters used in our experiments lie well within the ranges encountered by the animal in its natural environment. □

Received 25 August 1995; accepted 12 January 1996.

1. Schiff, S. J. *et al.* *Nature* **370**, 615–620 (1994).
2. Ruelle, D. *Physics Today* **47**, 24–30 (1994).
3. Ruelle, D. *Proc. R. Soc. Lond. A* **427**, 241–251 (1990).
4. Cvitanovic, P. *Physica D* **51**, 138–156 (1991).
5. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. *Physica D* **58**, 77–94 (1992).
6. Chang, T., Schiff, S. J., Sauer, T., Gossard, J.-P. & Burke, R. E. *Biophys. J.* **67**, 671–683 (1994).
7. Scott, D. A. & Schiff, S. J. *Biophys. J.* **69**, 1748–1757 (1995).
8. Kaplan, D. *Physica D* **73**, 38–48 (1994).
9. Cvitanovic, P. *Phys. Rev. Lett.* **61**, 2729–2732 (1988).
10. Pierson, D. & Moss, F. *Phys. Rev. Lett.* **75**, 2124–2127 (1995).
11. Ruelle, D. & Takens, F. *Commun. math. Phys.* **20**, 167–192 (1971).
12. Ruelle, D. *La Recherche* **108**, 132–246 (1980).
13. Ruelle, D. in *8th Int. Congress on Math. Phys.* (eds Mebkhout, M. & Senior, R.) 273–282 (World Scientific, Singapore, 1987).
14. May, R. M. *Nature* **261**, 459–467 (1976).
15. May, R. M. *Ann. N.Y. Acad. Sci.* **316**, 517–529 (1979).
16. Proc. 2nd Workshop on Measures of Complexity and Chaos (eds Abraham, N. B., Albano, A. M., Passamante, A. P., Rapp, P. E. & Gilmore, R.) *Int. J. Bifurc. Chaos* **3**, 485–490 (1993).
17. Grassberger, P. & Procaccia, I. *Physica D* **9**, 189–208 (1983).
18. Wolf, A., Swift, J. B., Swinney, H. L. & Vassano, J. A. *Physica D* **16**, 285–317 (1985).
19. Sugihara, G. & May, R. M. *Nature* **344**, 734–741 (1990).
20. Garfinkel, A., Spano, M. L., Ditto, W. L. & Weiss, J. N. *Science* **257**, 1230–1235 (1992).
21. Christini, D. J. & Collins, J. J. *Phys. Rev. Lett.* **75**, 2782–2785 (1995).
22. Kennedy, D. J. *gen. Physiol.* **46**, 551–572 (1963).
23. Wilkens, L. A. *Comp. Biochem. Physiol.* **91**, 61–68 (1988).
24. Wilkens, L. A. & Douglass, J. K. *J. exp. Biol.* **189**, 263–272 (1994).
25. Hayashi, H. & Ishizuka, A. *J. theor. Biol.* **156**, 269–291 (1992).
26. Bevington, P. R. *Data Reduction and Error Analysis* 48–49 (McGraw-Hill, New York, 1969).
27. Ditto, W. L., Rauseo, S. N. & Spano, M. L. *Phys. Rev. Lett.* **65**, 3211–3214 (1990).
28. Hunt, E. R. *Phys. Rev. Lett.* **67**, 1953–1955 (1991).
29. Roy, R., Murphy, T. W., Maier, T. D. & Gills, Z. *Phys. Rev. Lett.* **68**, 1259–1262 (1992).
30. Petrov, V., Gaspar, V., Masere, J. & Showalter, K. *Nature* **361**, 240–243 (1993).
31. Rollins, R. W., Parnananda, P. & Shepard, P. *Phys. Rev. E* **47**, R780–R784 (1993).
32. Artuso, R., Aurell, E. & Cvitanovic, P. *Nonlinearity* **3**, 325–360 (1990).
33. Artuso, R., Aurell, E. & Cvitanovic, P. *Nonlinearity* **3**, 361–395 (1990).
34. Shinbrot, R., Grebogi, C., Ott, E. & Yorke, J. A. *Nature* **363**, 411–417 (1993).
35. Moon, F. C. *Chaotic Vibrations* (Wiley, New York, 1987).
36. Strogatz, S. H. *Nonlinear Dynamics and Chaos Ch. 12* (Addison-Wesley, Reading, 1994).
37. Herman, H. T. & Olsen, R. E. *J. gen. Physiol.* **51**, 534–551 (1968).

ACKNOWLEDGEMENTS. We thank L. Wilkens, W. Ditto, M. Spano, D. Kaplan, M. Ding and L. Pecora for advice and discussions. This work was supported by the US Office of Naval Research.

Control of community growth and export production by upwelled iron in the equatorial Pacific Ocean

Kenneth H. Coale*, Steve E. Fitzwater*, R. Michael Gordon*, Kenneth S. Johnson*† & Richard T. Barber‡

* Moss Landing Marine Laboratories, PO Box 450, Moss Landing, California 95039-0450, USA

† Monterey Bay Aquarium Research Institute, PO Box 628, Moss Landing, California 95039-0628, USA

‡ Duke University, School of the Environment, Marine Laboratory, 135 Duke Marine Laboratory Road, Beaufort, North Carolina 28516-9721, USA

THE 'iron hypothesis'^{1,2} states that phytoplankton growth and biomass are limited by low concentrations of available iron in large regions of the world's oceans where other plant nutrients are abundant. Such limitation has been demonstrated by experiments in which iron has been added to both enclosed and *in situ* (un-enclosed) phytoplankton populations^{2–6}. A corollary of the iron hypothesis is that most 'new' iron is supplied by atmospheric deposition^{7,8}, and it has been suggested that changes in the deposition rates of iron-bearing dust have led to changes in biological productivity and, consequently, global climate⁷. Here we report surface-water measurements in the equatorial Pacific Ocean which show that the main iron source to equatorial waters at 140° W is from upwelling waters. Shipboard *in vitro* experiments indicate that sub-nanomolar increases in iron concentrations can cause substantial increases in carbon export to deeper waters in this region. These findings demonstrate that equatorial biological production is controlled not solely by atmospheric iron deposition, but also by processes which influence the rate of upwelling and the iron concentration in upwelled water.

Vertical profiles of iron and aluminum concentrations and primary production were determined at a number of stations between 3° S and 9° N along 140° W as part of the US Office of Naval Research FeLine programme and the Joint Global Ocean Flux Study, EqPac programme during 1990 and 1992. Samples were collected and analysed using trace-metal-clean methods developed and modified in our laboratory^{9,10}.

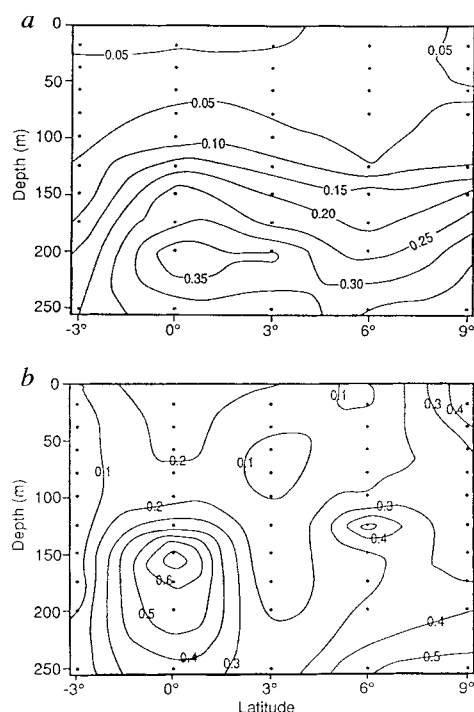


FIG. 1 The meridional distribution of dissolved iron (nmol kg^{-1} ; a) and particulate aluminum (nmol kg^{-1} ; b) along 140° W from 3° S to 9° N to a depth of 250 m (the depth range over which upwelling may occur). Circles represent sample latitude and depth. Iron concentrations were determined on vertical profiles taken during the FeLine cruise in July 1990. Samples were collected using Teflon-coated 30-l Go-Flo bottles suspended on Kevlar line. Sea water was pressure filtered from these bottles through 142-mm diameter. Nucleopore polycarbonate filters (0.4 μm pore size) inside a positive-pressure clean-air van. Analysis of dissolved iron used APDC/DDDC (1-pyrrolidinedithiocarbionic acid (APDC) ammonium salt; diethyldithiocarbionic acid (DDDC), diethylammonium salt) extraction into chloroform, modified to reduce iron blanks⁹. Surface-water iron concentrations are depleted, and are $\sim 0.05 \text{ nM}$ or less in the upper 100 m. At the Equator, isopleths of iron shoal in response to both strong vertical advection and a strong iron source in, and just below, the equatorial undercurrent. The depression of these isopleths at 3° N and 3° S corresponds to regions commonly associated with downwelling.

The meridional and vertical distribution of dissolved iron over this section is shown (as contours) in Fig. 1a. As atmospheric deposition of aerosol particles in the equatorial Pacific is low¹¹, the co-occurrence of high iron and particulate aluminum (Fig. 1b) along this transect indicate a terrigenous source of iron at the Equator, presumably originating in the western Pacific (New Guinea Platform¹²) and advecting eastwards in the equatorial undercurrent.

Recent surveys of the equatorial west Pacific, and source regions of the equatorial undercurrent, have revealed shallow, epithermal venting and iron-rich metalliferous sediments at a variety of locations^{13–15}. It is possible that shallow hydrothermal venting and benthic fluxes from these shelf sediments are major sources of this undercurrent iron.

The vertical distribution of iron concentrations at the Equator was combined with estimates of upwelling and vertical advection to calculate the flux of iron into the euphotic zone (Table 1). This calculation indicates that most (~90%) of the iron supply to the equatorial Pacific is from upwelling of iron-rich undercurrent waters and not from atmospheric deposition as previously thought^{7,8}. This flux is summarized and presented together with estimates of atmospheric deposition in Table 1.

Using a molar carbon:iron ratio for open-ocean phytoplankton of 167,000:1 (ref. 16) we calculate that this flux of iron could support a new primary production rate of 20 mmol C m⁻² d⁻¹. This is in good agreement with values of new production estimated by other means this area, which range from 5 to 30 mmol C m⁻² d⁻¹ (ref. 17). Inputs of new iron can only support enough production to draw down ~20% of the upwelled nitrate (Table 1). This observation, together with the high surface-water concentrations of nitrate (5–10 µM) strongly indicate that nitrate does not limit production in the equatorial Pacific. Because other trace nutrients such as Zn, Co and Ni did not show elevated concentrations in the undercurrent waters and are not thought to be biolimiting, we conclude that the supply of iron to the surface waters of the equatorial Pacific, primarily from upwelling, controls potential export production from this region. The supply of iron via upwel-

ling of terrigenous or hydrothermal inputs will decouple equatorial production from atmospheric deposition processes.

Such control requires that community production responds to upwelled iron, but upwelling contributes only picomolar amounts of iron to the surface waters. To examine the threshold of limitation, several enrichment experiments were performed as part of the FeLine and EqPac programmes, the results of which are presented in Fig. 2. For each experiment, iron in the control treatment was measured. In some cases, the initial controls showed elevated levels of iron as a result of processing. These cases were treated as iron additions and the results were evaluated at the measured concentration of iron.

Because the carbon:chlorophyll ratio changes on addition of iron (it drops as chlorophyll is preferentially synthesized⁴), and there is a large detrital component to particulate organic carbon (POC), community growth rates in all of these experiments were determined by the uptake of nitrate relative to initial particulate organic nitrogen (PON) over time. We assume that the decrease of dissolved nitrate is balanced by an increase in PON. This assumption has been substantiated by the observation that increases in dissolved organic nitrogen (DON) and ammonia are negligible with respect to the decrease in nitrate. These community growth rates were measured after the initial lag period (during a time of rapid exponential growth) and were converted to doubling rates¹⁸ (U in Fig. 2a and c). The results are plotted against iron concentration and are shown in Fig. 2. Although all phytoplankton taxa responded, diatoms exhibited the greatest response¹⁹. The iron-response curve (Fig. 2a) was derived using a nonlinear curve fitted to the Michaelis–Menten equation on untransformed data.

The results indicate a half-saturation constant of 0.12 nM with a maximum doubling of 0.99 per day. The doubling rate saturates at ~1 nM and does not increase with iron additions of up to 1 µM. With natural surface-water iron concentrations of less than 0.03 nM, we conclude that the iron-limited community could double its growth rate with the addition of only 0.1 nM Fe.

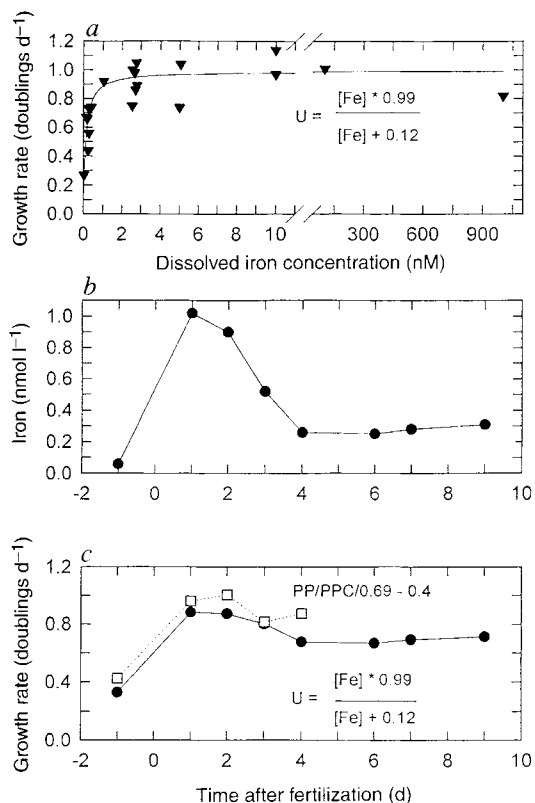


FIG. 2 a, Growth rate, U , (doublings per day) plotted against dissolved iron concentration. Each point represents the results of a single enrichment experiment. For each experiment, unfiltered sea water was collected using Teflon coated 30-l Go-Flo bottles (deployed on a Kevlar hydrowire and tripped with a Teflon messenger) and aliquoted sequentially into 2-l (EqPac studies⁹), or 20-l (FeLine studies^{4,5}) acid-cleaned polycarbonate bottles, under class 100 clean-room conditions on board ship. The bottles were then either spiked with iron (iron chloride or iron sulphate) or left as controls and incubated in shipboard incubators at surface water temperatures for 5–7 d. In the case of the controls, initial iron concentrations were measured. Treatments and controls were sampled every day for chlorophyll and nutrients. Maximum growth rates were calculated, during the exponential growth phase, using initial PON and the time-rate depletion of nitrate as a proxy for the increase in PON¹⁸. As such, these growth rates represent net community growth rates and do not reflect the response of any single group or taxa, although diatoms dominated the increase in chlorophyll. b, Iron concentrations measured within the enriched patch in discrete samples during the IronEx-I experiment² as a function of time. Ambient iron concentrations of 0.05 nM were increased to ~4 nM with the addition of acidic iron sulphate. This figure indicates that the added iron was rapidly lost from the system. c, The Michaelis–Menten equation derived in a together with the measured dissolved iron concentrations (b) were used to calculate the community growth rate (filled circles). These values agree closely with the growth rate (PP) measured using H¹⁴CO₃⁻ uptake together with epifluorescent microscopic determinations of phytoplankton biomass (PPC) for each day before subduction of the patch². The measured doubling rates were adjusted by 0.4 to compensate for respiration. This factor represents the respiration rates recently reported for the equatorial Pacific in this region¹⁷.

Price and co-workers²⁰ report similar findings in a study of iron enhancement of nitrate uptake rates, also in the equatorial Pacific. They found that the presence of ammonium would inhibit nitrate uptake in the small phytoplankton that did not appear to be iron-limited, yet not enough iron was present to allow larger phytoplankton to take up nitrate. Nitrate uptake was enhanced, however, by the larger phytoplankton in the presence of added iron. They concluded that the uptake of nitrate by the community (therefore new production) was ultimately limited by iron availability but that phytoplankton biomass was controlled by the presence of grazers. In addition, Takeda and Obata²¹ report the results of bottle enrichment experiments which also showed that equatorial phytoplankton can respond to subnanomolar additions of iron. Our measurements of extremely low surface-water concentrations of dissolved iron in the equatorial Pacific have been substantiated by others using a variety of independent techniques^{21,22}.

It has been argued that bottle enrichment experiments, because of containment artefacts and an undersampling of the population, do not accurately represent the community response. We have therefore used our experimentally derived Michaelis-Menten relationship to calculate community growth rates during the recent unenclosed enrichment experiment in the equatorial Pacific². Figure 2b depicts the variation over time of the iron concentration in the experimental patch, indicating the rapid decrease from an initial iron concentration of ~ 4 nM. Iron values from Fig. 2b were then used to calculate community

TABLE 1 Calculated fluxes of iron, nitrate and carbon

	Flux (nmol m ⁻² d ⁻¹)	Potential new production (mmol C m ⁻² d ⁻¹)
Iron source		
Upwelling	120	20
Diffusion	0.20	0.03
Atmospheric	5–25	0.8–4.2
Total	125.2–145.2	20.8–24.2
Nitrate source		
Upwelling	15.48×10^6	103
Diffusion	0.008×10^6	0.05
Atmospheric	0.005×10^6	0.03
Total	15×10^6	103.1

The supply of iron to the equatorial Pacific at 140°W during the EqPac and FeLine cruises were calculated using the simple one dimensional model: $J_{Fe} = w[Fe]_d + K_z dFe_d/dz$, where J_{Fe} is the vertical flux of iron (nmol m⁻² d⁻¹), $[Fe]_d$ is the concentration of dissolved iron (nmol m⁻³), dFe_d/dz is the gradient in dissolved iron with depth (nmol m⁻⁴), w is the upwelling velocity (m d⁻¹) and K_z is the vertical diffusivity constant (m² d⁻¹). Equatorial upwelling rates, diffusivity constants and atmospheric deposition rates were taken from the literature^{11,27–29}. Iron concentrations and gradients were derived from a composite profile of all the EqPac and FeLine dissolved iron data ($Fe(nM) = -0.0357 + 0.00623Z - 1.79 \times 10^{-4} Z^2 + 2.09 \times 10^{-6} Z^3 - 1.06 \times 10^{-8} Z^4 + 2.40 \times 10^{-11} Z^5 - 2.02 \times 10^{-14} Z^6$, $r^2 = 0.90$). Iron data are available: see Supplementary Information. Surface-water concentrations below our detection limit (2 times the standard deviation of the blank analyses, 0.03 nM) were taken to be 0.026 nM. Nitrate fluxes are calculated in a similar way. The fluxes (via upwelling, diffusion and atmospheric deposition) are presented, together with the calculated potential new production that these fluxes could support. The molar carbon:iron ratio of 167,000 was used to calculate potential production¹⁶. Although there is considerable variation in this reported ratio, this value was determined for oceanic phytoplankton and is mid-range between reported values. All fluxes were calculated as inputs to a 120-m-thick euphotic zone (0.1% light level). Iron flux via upwelling was calculated as the product of the upwelling velocity at 120 m ($w = 1.23$ m d⁻¹) and the iron concentration at that depth. Iron flux via diffusion was calculated using the derivative of the composite curve (above) at 120 m times the diffusion coefficient ($K_z = 0.086$ m² d⁻¹). Iron flux limits new production to $\sim 20\%$ of the potential based on nitrate.

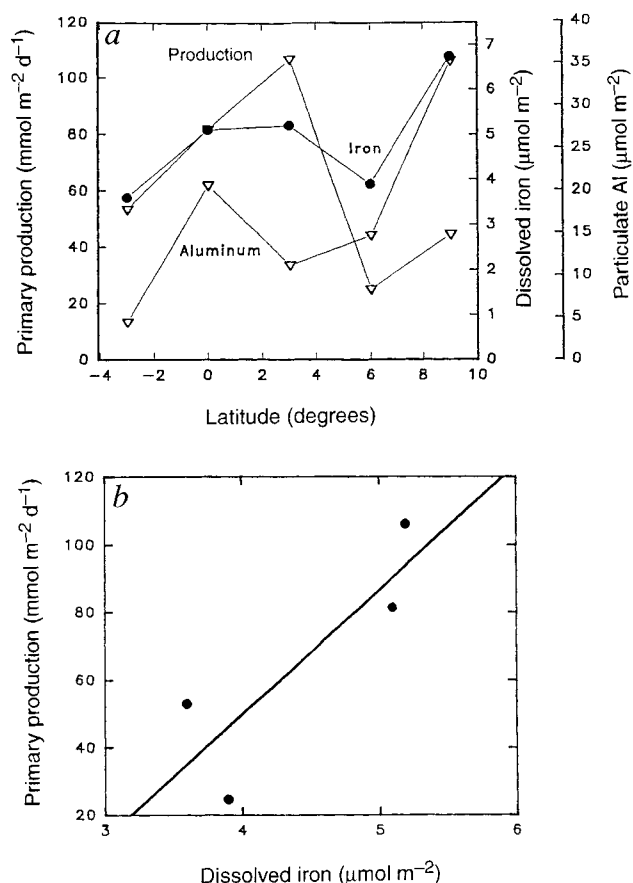


FIG. 3 a, The meridional distribution of dissolved iron, particulate aluminum and primary production integrated over the euphotic zone during the FeLine cruise of June 1990. This transect indicates populations limited by nitrate at the northern stations and iron at the southern stations. The distribution of particulate aluminum is an indicator of the atmospheric deposition of terrigenous material except at the Equator where the undercurrent is responsible for this signal. b, Integrated primary production plotted against integrated iron, showing a strong correlation ($r^2 = 0.73$) for the stations 3°S, 0°, 3°N and 6°N along 140°W, indicating iron limitation throughout this region. The station at 9°N, in spite of the relatively high iron concentration, was nitrate limited ($[NO_3^-] < 0.1 \mu M$) and therefore does not follow this relationship. Shipboard enrichment experiments indicate that other metals (for example, Zn, Co, Mn) are not limiting^{6,23,30}.

doubling rates (Fig. 2c, filled circles). These agree closely with community doubling rates measured during this period (squares) indicating that the growth response curve obtained in these bottle experiments (Fig. 2a) is environmentally robust. Because surface-water dissolved iron concentrations in the open ocean are generally 0.05 nM or less, and midwater (1,000–2,000 m) iron concentrations do not exceed 0.8 nM, we hypothesize that this relationship can be extrapolated to other high-nitrogen low-chlorophyll environments.

Measurements of dissolved iron and primary production in a meridional transect from 3°S to 9°N along 140°W provide further field observations that support these findings. Primary production and dissolved iron, integrated to the 0.1% light level for this meridional transect, are plotted together in Fig. 3a. There is a general decrease in dissolved iron concentration from north to south, with an intermediate maximum associated with the region of equatorial upwelling. Except for the local maximum near the Equator, where upwelling dominates the input of iron to the euphotic zone, this trend follows the pattern of atmospheric deposition¹¹. The distribution of particulate aluminum also follows this same pattern (Fig. 3a).

Although the data set is limited, there is a correlation between

depth-integrated primary production and the integrated dissolved-iron concentration along this section (Fig. 3b), except at 9°N where the nitrate concentrations are extremely low ($< 0.1 \mu\text{M}$). This finding is consistent with measurements of photon energy conversion efficiency in phytoplankton made by fast-repetition-rate-fluorometry²³ and of photoconversion efficiency as determined by measurements of photosynthesis versus irradiance²⁴.

We would expect that integrated primary productivity would be correlated with the flux of iron to the euphotic zone because both reflect rate processes, yet no clear relationship between the flux of iron and its concentration in the euphotic zone has been demonstrated in this region. Measured concentrations of iron will reflect both the rate of delivery and the rate of removal. If removal is proportional to production^{25,26}, then the observed correlation requires that the concentrations of iron and aluminum co-vary with flux in this region. Although the EqPac data for other years do not fall on the line generated by the FeLine transect of 1990, the general trend is conserved: higher integrated primary production correlates with higher integrated iron concentrations.

The data reported here implicate equatorial upwelling of iron as a controlling factor in sustaining elevated levels of production in the equatorial Pacific. They also demonstrate substantial enhancement of phytoplankton growth in these waters with only picomolar additions or iron, show that natural populations of plankton are growing well below their half saturation constant for iron, and demonstrate that the meridional variation in primary productivity is correlated with the distribution of dissolved iron.

Variations in the source(s) of iron to these undercurrent waters and the equatorial upwelling rate, the main driving force for iron transport into the surface waters, must have a direct effect on the variability of equatorial productivity and carbon export from this system. □

Received 13 January 1995; accepted 23 January 1996.

- Geider, R. J. & LaRoche, J. *Photosynth. Res.* **39**, 275–301 (1994).
- Martin, J. H. et al. *Nature* **371**, 123–129 (1994).
- Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
- Coale, K. H. *Limnol. Oceanogr.* **36**, 1851–1864 (1991).
- Johnson, K. S., Coale, K. H., Elrod, B. A. & Tindale, N. W. *Mar. Chem.* **46**, 319–334 (1994).
- Kolber, Z. S. et al. *Nature* **371**, 145–149 (1994).
- Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).
- Leinen, M. & Murray, J. W. (eds) *A Rationale and Plan for U.S. Joint Global Ocean Flux Studies in the Central Equatorial Pacific, Problem II, Hypothesis 2*. (USJGOFS Planning & Coordination Office, Woods Hole Oceanographic Instn, Woods Hole, MA 1990).
- Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. *Analytica chim. Acta* **105**, 233–245 (1979).
- Eggiman, D. W. & Betzer, P. R. *Analyt. Chem.* **48**, 866–890 (1976).
- Duce, R. A. & Tindale, N. W. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
- Tsuchiya, M., Lukas, R., Fine, R. A., Firing, E. & Lindstrom, E. *Prog. Oceanogr.* **23**, 101–147 (1989).
- Herzig, P. et al. *Eos* **75**, 513–515 (1994).
- Sudankov, S. M., Davydov, M. P., Bazelyan, V. L. & Tarasov, V. G. in *Hydrothermal Vents and Processes* (eds Parson, L. M., Walker, C. L. & Dixon, D. R.) 249–255 (Spec. Publ. No. 87, Geological Soc., Bath, UK, 1995).
- Exon, N. F. & Cronan, D. S. *Mar. Geol.* **52**, M43–M52 (1983).
- Morel, F. M. M., Hudson, R. J. M. & Price, N. M. *Limnol. Oceanogr.* **36**, 1742–1755 (1991).
- Murray, J. W., Barber, R. T., Roman, M. R., Bacon, M. P. & Feely, R. A. *Science* **266**, 58–65 (1994).
- Banse, K. *Limnol. Oceanogr.* **36**, 1886–1878 (1991).
- Fitzwater, S. E., Coale, K. H., Gordon, R. M., Johnson, K. S. & Ondrusek, M. E. *Deep-Sea Res.* (in the press).
- Price, N. M., Ahner, B. A. & Morel, F. M. M. *Limnol. Oceanogr.* **39**, 520–539 (1994).
- Takeda, S. & Obata, H. *Mar. Chem.* **50**, 219–241 (1995).
- Rue, E. & Bruland, K. W. *Eos* **26**, OS154 (1996).
- Green, R. M., Kolber, Z. S., Swift, D. G., Tindale, N. W. & Falkowski, P. G. *Limnol. Oceanogr.* **39**, 1060–1074 (1994).
- Lindley, S., Bidigare, R. & Barber, R. T. *Deep-Sea Res.* (in the press).
- Coale, K. H. & Bruland, K. W. *Limnol. Oceanogr.* **32**, 189–200 (1987).
- Bruland, K. W. & Coale, K. H. in *Dynamic Processes in the Chemistry of the Upper Ocean*. (eds Burton, J. D., Brewer, P. G. & Chesselet, R.) 159–172 (Plenum, New York, 1986).
- Gargett, A. E. *Limnol. Oceanogr.* **36**, 1527–1545 (1991).
- Bryden, H. L. & Brady, E. C. *J. phys. Oceanogr.* **15**, 1255–1273 (1985).
- Peters, H., Gregg, M. C. & Toole, J. M. *J. geophys. Res.* **93**, 1199–1218 (1988).
- Sanderson, M. & Bruland, K. *Eos* **76**, OS177 (1996).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank our colleagues S. Tanner, C. Hunter, G. Elrod and J. Nowicki for their assistance. Our appreciation is also extended to the crew and officers of the R/V Thompson and R/V Wecoma. This work was supported by the US National Science Foundation and the US Office of Naval Research.

Constraints on early Earth differentiation from hafnium and neodymium isotopes

J. D. Vervoort*, P. J. Patchett*, G. E. Gehrels* & A. P. Nutman†

* Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

† Research School of Earth Sciences, Australian National University, Canberra ACT 0200, Australia

INFERENCES about the early evolution of the Earth's crust and mantle have come largely from the study of isotope systematics—in particular, those of neodymium^{1–5}. Neodymium isotope data from the oldest preserved rocks have been interpreted^{6–8} as reflecting early large-scale chemical depletion of the mantle (presumably resulting from the extraction of continental crust), but these data have remained controversial, in view of the potential for disturbances to the samarium–neodymium system during these rocks' long history^{9–11}. Here we provide an independent evaluation of the Nd isotope compositions of ten early Archaean (3.6–3.8 Gyr old) gneisses, by investigating the hafnium isotope systematics of zircons from these rocks. The Hf data are consistent with the Nd record in indicating early depletion of the mantle, but fail to verify the scale and variability of this depletion. We conclude that Nd isotopes of early Archaean gneisses do not faithfully record isotopic variations in the early Earth, and therefore that these data need to be examined more critically before they can be used to constrain the early history of crust–mantle differentiation.

The neodymium isotope record of juvenile, mantle-derived rocks is quite well documented^{12,13}. In general terms, the record of the highest ϵ_{Nd} values^{4,13,14} can be described in three main segments (Fig. 1): (1) a nearly linear increase from about 2.5 Gyr ago to the present; (2) an essentially static period from 3.6 to 2.5 Gyr; and (3) a positive 'spike' in initial ϵ_{Nd} from 4.0 to 3.6 Gyr. The increasingly positive trend in initial ϵ_{Nd} values of mantle-derived rocks from 2.5 Gyr to today^{15,16} can be explained by the extraction of Archaean crust from the mantle, causing incompatible-element depletion and, with time, increasingly positive ϵ_{Nd} . The existence of a widespread depleted mantle reservoir is justified by extensive preserved crust from 3.1–2.6 Gyr ago.

The Archaean record, however, is problematical because initial ϵ_{Nd} values of demonstrably juvenile crust are predominantly positive and average about +2 and +3 for the period 3.6 to 2.5 Gyr (Fig. 1). These positive initial ϵ_{Nd} values are more difficult to explain by a simplistic depleted mantle reservoir because very little continental crust from 4.0 to 3.1 Gyr is preserved. Identified pre-3.1 Gyr crustal terranes account for much less than 5% of continental crust. This flat part of the Nd evolutionary curve may reflect mixing of mantle reservoirs that have undergone different amounts of depletion (or enrichment)⁴. Interpretation of the Nd record is further complicated by the apparent positive spike in initial ϵ_{Nd} values as high as +4.5 for 4.0–3.6 Gyr gneisses^{4,6–8}. A complementary negative spike as low as $\epsilon_{\text{Nd}} = -4$ has also been reported for early Archaean gneisses^{4,8}. These Nd data appear to necessitate differentiation of the Earth into crust and mantle very early in its history. Early and widespread terrestrial differentiation is consistent with some Pb-isotope evidence¹⁷, but is at odds with evidence from coupled ^{146}Sm – ^{142}Nd and ^{147}Sm – ^{143}Nd systematics. The lack of persistent detectable $^{142}\text{Nd}/^{144}\text{Nd}$ anomalies in rocks with positive initial ϵ_{Nd} values^{18,19} indicates that the radiogenic $^{143}\text{Nd}/^{144}\text{Nd}$ compositions were not produced by widespread mantle differentiation before 4.3 Gyr. If all these data are valid, then extreme Sm/Nd fractionations would have to have occurred in the mantle by 4.3–4.2 Gyr to generate the large range in $^{143}\text{Nd}/^{144}\text{Nd}$ compositions by 4.0–3.8 Gyr (refs 3–5).

The validity of the $^{143}\text{Nd}/^{144}\text{Nd}$ isotope record of early Archaean gneisses has been questioned, however, because all early Archaean Nd data come from whole-rock analyses of highly (and multiply) metamorphosed gneisses. Only slight changes in Sm/Nd are required to modify the apparent initial $^{143}\text{Nd}/^{144}\text{Nd}$ in these old rocks¹⁰. Some have suggested that the highly positive ϵ_{Nd} values of many early Archaean gneisses are artefacts of disturbances in the Sm–Nd isotope system^{9–11}. These problems, real or imagined, have prevented Nd isotope data for early Archaean gneisses from gaining complete acceptance and becoming a fixed point in discussions of early Earth evolution²⁰.

Hafnium isotopes can provide an independent evaluation of the neodymium data, because in the mantle environment the geochemical behaviour of Lu/Hf parallels that of Sm/Nd. This results in the correlation of $^{176}\text{Hf}/^{177}\text{Hf}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ in nearly all mantle (and crustal) rocks. Hf and Nd correlation in mantle rocks was first recognized in ocean island basalts (OIBs)^{21–23}, and was originally approximated by the relation $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$. This

relationship is also predicted by partition coefficients of Hf and the rare-earth elements for probable mantle residues^{24,25}. Although the Hf–Nd relationship for the OIB source region is now thought by some²⁶ to be more like $\epsilon_{\text{Hf}} = 1.4\epsilon_{\text{Nd}}$, the record of juvenile crust^{15,27–31} throughout the geological record is more consistent with the $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$ relationship (Fig. 2). These juvenile rocks are significant for our purposes as they most probably represent new additions to continental crust from the mantle.

Zircons are ideally suited for determining initial Hf isotope compositions because (1) they have abundant Hf and very low Lu/Hf ratios; (2) crystallization ages can be determined directly from them; (3) they are resistant to Lu/Hf disturbance; and (4) disturbances, if they do occur, can be monitored by U–Pb systematics. The abundant Hf and low $^{176}\text{Lu}/^{177}\text{Hf}$ (usually < 0.0005) in zircon is important because it means that $^{176}\text{Hf}/^{177}\text{Hf}$ is insensitive to changes in Lu/Hf. The only way to significantly alter $^{176}\text{Hf}/^{177}\text{Hf}$ of zircon is to exchange Hf with more radiogenic minerals in the rock, or to add younger radiogenic Hf as an overgrowth^{22,32}. In the

TABLE 1 Hf and Nd isotopes in early Archaean Greenland gneisses

Sample	Size fraction (μm)	A	Age (Gyr)	$\frac{^{176}\text{Hf}}{^{177}\text{Hf}}$	$\frac{^{176}\text{Lu}}{^{177}\text{Hf}}$	$\epsilon_{\text{Hf}}(T)$ §	$\frac{^{143}\text{Nd}}{^{144}\text{Nd}}$	$\frac{^{147}\text{Sm}}{^{144}\text{Nd}}$	Sm‡ (p.p.m.)	Nd‡ (p.p.m.)	$\epsilon_{\text{Nd}}(T)$
Amitsoq gneisses, Isukasia region											
GGU-248226	(bulk)		3.73¶	0.280433 ± 12	0.00007	3.6 ± 0.4	0.511593	0.1512	7.5	29.9	1.5
	(80–125)			0.280445 ± 40	0.00016	3.8 ± 1.4					
	(20–45)			0.280434 ± 12	0.00015	3.5 ± 0.4					
GGU-248228	(bulk)		3.69¶, #	0.280463 ± 20	0.00011	3.7 ± 0.7	0.510736	0.1157	3.5	18.3	1.5
	(80–125)	A		0.280461 ± 16	0.00062	2.3 ± 0.6					
	(80–125)			0.280461 ± 21	0.00001	3.9 ± 0.7					
GGU-248245	bulk		3.70¶	0.280377 ± 8	0.00026	0.5 ± 0.3	0.510106	0.0893	5.4	36.9	1.9
	(45–80)	A		0.280406 ± 13	0.00051	0.8 ± 0.3					
GGU-248250	(80–125)		3.70¶	0.280417 ± 15	0.00011	2.3 ± 0.5	0.510283	0.0975	1.9	11.6	1.5
	(45–80)			0.280428 ± 12	0.00017	2.6 ± 0.4					
	(45–80)			0.280418 ± 13	0.00003	2.6 ± 0.5					
91-680	(> 80)	A	3.81 ⁴	0.280388 ± 8	0.00046	3.0 ± 0.3					3.5 ^{4, ☆}
90-410	(bulk)	A	3.81 ⁴	0.280388 ± 13	0.00016	3.5 ± 0.5					4.0 ^{4, ☆}
	(80–125)	A		0.280396 ± 11	0.00032	3.7 ± 0.4					4.5 ^{4, ☆}
	(> 125)	A		0.280358 ± 10	0.00032	2.4 ± 0.4					4.2 ^{4, ☆}
	(> 125)	A		0.280382 ± 7	0.00028	3.1 ± 0.3					
Isua supracrustals, Isukasia region											
248203	(45–80)		3.80¶	0.280406 ± 17	0.00023	4.1 ± 0.6	0.510156	0.0928	3.8	24.5	2.6
Amitsoq gneisses, Nuuk region											
GGU-110999	(80–125)		3.82 ³³	0.280388 ± 10	0.00007	4.4 ± 0.4	0.510551	0.1088	3.2	17.5	2.7
	(80–125)	A		0.280415 ± 12	0.00034	4.6 ± 0.3					
GGU-155817	(> 125)		3.65¶	0.280490 ± 23	0.00013	3.6 ± 0.8	0.510795	0.1163	8.7	45.3	2.0
GGU-155820	(> 125)		3.64¶	0.280533 ± 8	0.00010	5.0 ± 0.3	0.511159	0.1325	13.3	60.8	1.1
	(> 125)	A		0.280543 ± 14	0.00056	4.1 ± 0.5					
Flin-Flon area, Amisk Lake rhyolite											
Flin-Flon A			1.88 ³⁴	0.281904 ± 13	0.00021	10.9 ± 0.5	0.512368	0.1543	6.0	19.8	5.0

Approximately 1 mg of zircons were selected by removing non-zircon and altered, fractured, or composite grains, and dissolved (in concentrated hydrofluoric acid) in standard steel-jacketed Teflon digestion vessels. Following dissolution, aliquots were taken for U–Pb analysis (~1%) and Lu–Hf concentration (~20%). Hf analyses were made on both unabraded and abraded fractions. In addition to the U–Pb aliquots, single zircon grains (both abraded and unabraded) were also analysed for U–Pb. Sm–Nd analyses were performed on whole-rock powders from which the zircons were extracted. All isotope measurements were performed at the University of Arizona with the exception of U–Pb and Nd data from samples 91–680 and 90–410⁴.

* A in this column indicates abraded.

† Ratios normalized to $^{176}\text{Hf}/^{177}\text{Hf} = 0.7325$. JMC475 Hf standard gave $^{176}\text{Hf}/^{177}\text{Hf} = 0.282169 \pm 22$ (2 s.d. of the population; $n = 24$) for the JMC475 Hf standard during the study.

‡ 2 σ errors for Sm and Nd concentrations and $^{147}\text{Sm}/^{144}\text{Nd}$ are < 0.5%; 2 σ errors for $^{176}\text{Lu}/^{177}\text{Hf}$ are < 1.0%.

§ $\epsilon_{\text{Hf}} = [(\frac{^{176}\text{Hf}}{^{177}\text{Hf}})_{\text{sample}} / (\frac{^{176}\text{Hf}}{^{177}\text{Hf}})_{\text{CHUR}} - 1] \times 10^4$, where $(\frac{^{176}\text{Hf}}{^{177}\text{Hf}})_{\text{CHUR}} = 0.282830$. $\epsilon_{\text{Hf}}(T)$ is the ϵ_{Hf} value for a rock at the time (T) of its formation. The value of T used for the calculation of $\epsilon_{\text{Hf}}(T)$ is the age given in this table. The CHUR (chondritic reference) value was adjusted down compared to Patchett²² to reflect current values for $^{176}\text{Hf}/^{177}\text{Hf}$ in JMC475.

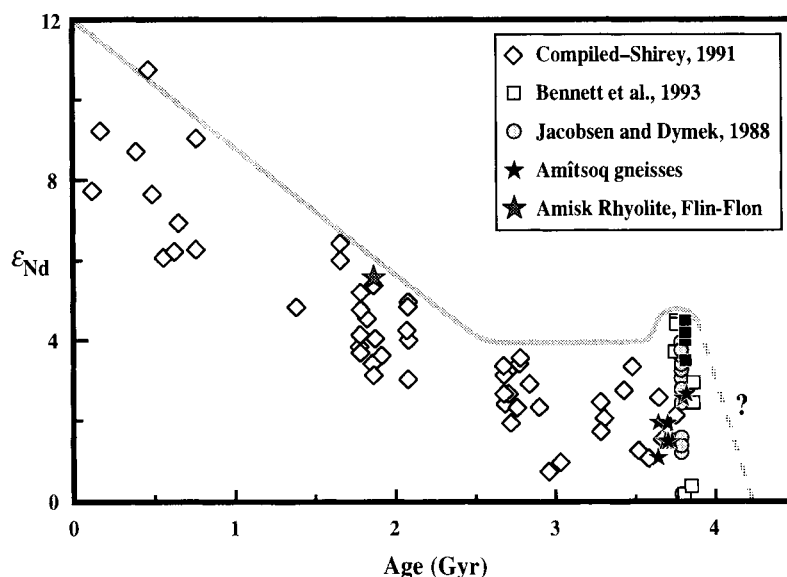
|| Ratios normalized to $^{146}\text{Nd}/^{144}\text{Nd} = 0.7219$. LaJolla Nd standard gave $^{143}\text{Nd}/^{144}\text{Nd} = 0.511870 \pm 11$ (two s.d. of the population; $n = 30$) during the study. All $^{143}\text{Nd}/^{144}\text{Nd}$ ratios were bias corrected to LaJolla $^{143}\text{Nd}/^{144}\text{Nd} = 0.511860$.

¶ U–Pb ages from G. Gehrels, unpublished data.

Also dated by ion microprobe: weighted mean $^{207}\text{Pb}/^{206}\text{Pb}$ age of $3,693 \pm 4$ Myr (A. Nutman, personal communication).

☆ Replicate analyses of separate whole-rock powders⁴ do not correspond to different zircon fractions.

FIG. 1 Upper range of initial ϵ_{Nd} values as a function of age. Shaded line represents approximate upper limit of range. Only the more positive initial ϵ_{Nd} values are shown. The ϵ_{Nd} notation (defined as $\epsilon_{\text{Nd}} = [({}^{143}\text{Nd}/{}^{144}\text{Nd})_{\text{sample}} / ({}^{143}\text{Nd}/{}^{144}\text{Nd})_{\text{CHUR}} - 1] \times 10^4$) is used by geochemists as a convenient shorthand for describing variations in ${}^{143}\text{Nd}/{}^{144}\text{Nd}$ relative to a chondritic reference (CHUR) composition—believed to represent a bulk Earth (undifferentiated) composition. Rocks with initial ϵ_{Nd} (compositions at the time of their formation) greater than zero, have been derived from a reservoir enriched in ${}^{143}\text{Nd}/{}^{144}\text{Nd}$ relative to CHUR and are interpreted to be produced by depletion of more incompatible elements during a mantle melting event. Compiled from Shirey¹³ (open diamond symbols). Other data are from Bennett *et al.*⁴ (open squares) and Jacobsen and Dymek¹⁴ (shaded circles). Filled squares are samples from Bennett *et al.*⁴, analysed in this study. Solid and shaded stars are other samples from Greenland and Flin-Flon, respectively, analysed in this study.



case of overgrowths or inherited cores, useful Hf information will be unattainable as the zircons will contain an indecipherable mixture of Hf isotope compositions. Fortunately, such cases should be easily detectable with integrated U–Pb and Hf isotopic work.

The Greenland samples analysed in this study have been subjected to multiple episodes of metamorphism in the period 3.8–3.6 Gyr, as shown by varying degrees of U–Pb complexity in the zircons: from simple and mostly concordant to highly discordant with prominent ancient Pb loss. In spite of the varying U–Pb complexities in the zircons and their diverse post-crystallization histories, the Hf isotope composition of different fractions (abraded and unabraded, different size fractions, varying discordance; Table 1^{4,33,34}) are identical for a given sample. If the zircons had either younger overgrowths or inherited cores, we would see differences in Hf isotope compositions in abraded and unabraded fractions, which we do not. These results are of great importance in demonstrating that the Hf isotope compositions have not been affected by processes that produced U–Pb disturbances.

The early Archaean gneisses that we have analysed for Hf and Nd (Table 1) range in age from 3.82 to 3.65 Gyr and are from the Nuuk and Isukasia (near the Isua supracrustal belt) areas of Greenland^{35–37}. Two samples were previously analysed by ion microprobe methods (SHRIMP)⁴. Conventional U–Pb analyses were performed on aliquots from the Hf dissolutions and on individual abraded and unabraded zircons (For U–Pb data, see Supplementary Information). The U–Pb work was done not only to constrain crystallization ages for determining initial ϵ_{Hf} and ϵ_{Nd} , but also to monitor disturbance to the U–Pb system in the zircons. Although precise ages are important for accurate determinations of both ϵ_{Hf} and ϵ_{Nd} values, the relation of ϵ_{Hf} to ϵ_{Nd} in these rocks is insensitive to age within a 200-Myr time frame.

Eight of the Greenland gneisses that we analysed have initial ϵ_{Nd} values that range from +1.1 to +2.7. These values are common in 3.6–3.73 Gyr rocks in several terranes worldwide¹³. However the database for single-component pre-3.73 Gyr rocks is restricted to Greenland and Labrador. This is because all the pre-3.73-Gyr rocks from other terranes (Canada, China, and Antarctica) are migmatites or banded gneisses^{38–40}. Of the pre-3.8-Gyr rocks from Greenland, we have analysed two samples with initial ϵ_{Nd} values of +3.5 to +4.5 (ref. 4) to represent the high end of ϵ_{Nd} values for the Earth's oldest preserved rocks^{4,7,8}.

The most significant results of this work are shown in Fig. 3. The initial ϵ_{Hf} values of the Greenland gneisses range from +0.5 to +5.0. Excluding sample 248245, which appears to be an outlier

(Fig. 3b), the initial ϵ_{Hf} values in these rocks are quite consistent (average +3.6; range, +2 to +5) and appear less variable than whole-rock ϵ_{Nd} values (Fig. 3b). The more uniform initial ϵ_{Hf} values in these rocks are remarkable considering the larger external reproducibility of Hf (approximately $\pm 0.8 \epsilon_{\text{Hf}}$ units) and the potential for diverse Hf compositions due to zircons with younger overgrowths or inherited cores. This consistency adds credibility to the Hf data, and further demonstrates that the zircon Hf systematics are not recording hypothetical unusual Lu/Hf fractionation in the history of the total rock. If the general relationship of $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$ approximates Hf–Nd isotope compositions of the early Archaean mantle, then initial ϵ_{Hf} values of +2 to +5 would appear to be consistent with initial ϵ_{Nd} values in the +1 to +1.5 range but not with ϵ_{Nd} values greater than +3.

Perhaps the most controversial samples are 91-680 and 90-410,

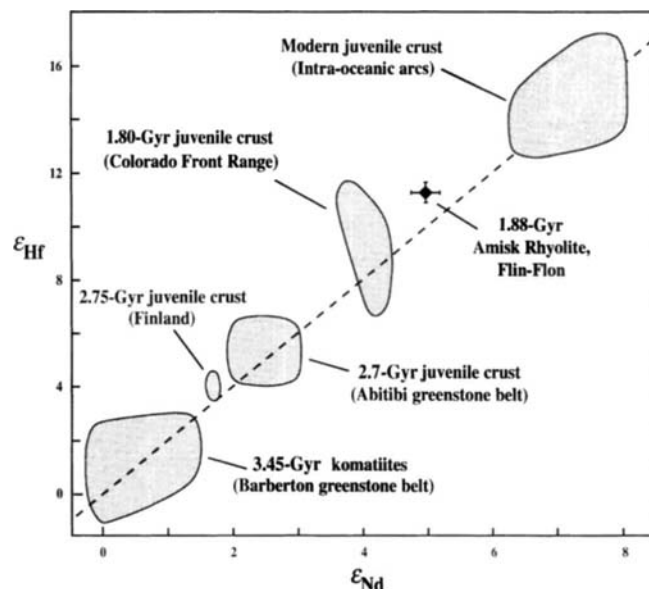
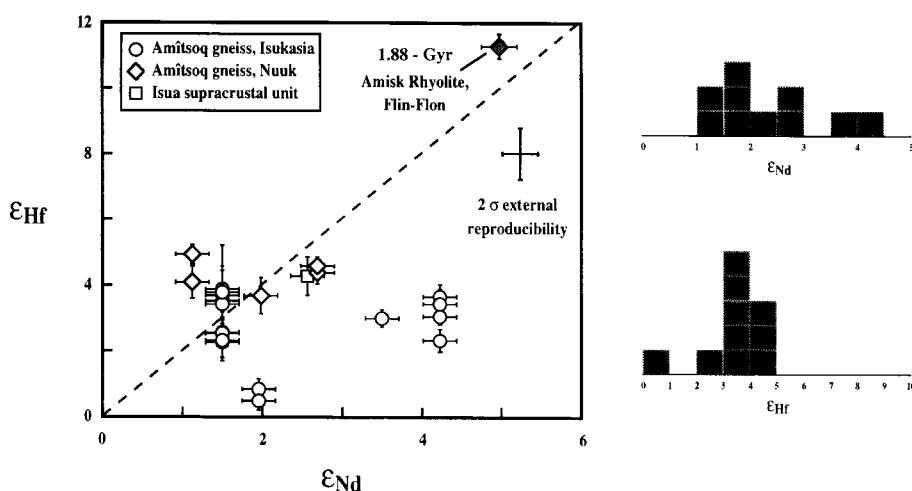


FIG. 2 Initial Hf and Nd isotope compositions of select juvenile crust of various ages: Barberton greenstone belt²⁷; 2.75-Gyr crust in Finland²⁸; Abitibi greenstone belt²⁹; Colorado Front Range^{12,30}; Flin-Flon volcanics, Trans-Hudson Orogen (Table 1); modern intra-oceanic arcs³¹. Dashed line represents $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$, the approximate relation expected between Hf and Nd in the early Archaean juvenile rocks.

FIG. 3 Left, initial ϵ_{Hf} and ϵ_{Nd} compositions of the Greenland gneisses. Dashed line corresponds to the relationship $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$ observed in virtually all juvenile crustal materials and which is predicted by partition coefficients. Initial ϵ_{Nd} values vary from +1.1 to +4.5. Hf isotopes, however, are not correlated with Nd and do not confirm the initial ϵ_{Nd} variations. This is especially evident for samples with ϵ_{Nd} values > 3. These rocks have initial ϵ_{Hf} of +2.4 to +3.7, much lower than the initial ϵ_{Hf} values of +6 to +10 expected from the $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$ relationship. Hf and Nd isotope composition of 1.88-Gyr rhyolite from the juvenile Flin-Flon belt shows the position of a well behaved sample with higher ϵ_{Nd} values. Right, histograms of ϵ_{Hf} and ϵ_{Nd} values, illustrating the more consistent isotopic signature of Hf compared to Nd.



for which whole-rock initial ϵ_{Nd} values of +3.5 to +4.5 have been reported⁴. We determined initial ϵ_{Hf} values for zircons from these rocks that range from +2.4 to +3.7. These ϵ_{Hf} values are well within the narrow range of values for the other Greenland gneisses with lower initial ϵ_{Nd} values and are much lower than ϵ_{Hf} values of +6 to +10 predicted by the $\epsilon_{\text{Hf}} = 2\epsilon_{\text{Nd}}$ relationship. The Hf data suggest that the high initial ϵ_{Nd} values calculated for these samples are artefacts of later metamorphic disturbances in the Sm–Nd system, and do not accurately record the Nd isotope composition of their source regions at 3.81 Gyr. Small changes in Sm/Nd are easy to justify given the high-grade polymetamorphic nature of these old gneisses¹⁰. However, the sample with the highest initial ϵ_{Nd} value (90-410) is the most pristine of any of our samples. It is a homogenous quartz diorite with igneous texture still preserved. The SHRIMP U–Pb age determinations on this sample are nearly all concordant⁴; the zircon grains have minimal ancient Pb loss and are devoid of both inherited cores or younger overgrowths. There is no justifiable way to dismiss this sample on the basis of obvious macroscopic, microscopic or cryptic U–Pb isotopic disturbances. If the Sm–Nd data from this sample are suspect, then we must be wary of Sm–Nd data from all such old, high-grade samples.

The Hf and Nd isotope data from the early Archaean Greenland gneisses portray somewhat different images of early Earth differentiation. On one hand, both Hf and Nd indicate that the mantle was depleted in incompatible elements (presumably due to extraction of crust from the mantle) very early in its history (4.3–4.2 Gyr). On the other hand, Hf isotopes fail to verify the extreme mantle depletion as indicated by the highly positive initial ϵ_{Nd} values. We cannot dismiss individual samples based on their Hf–Nd compositions, owing to the possibility of unusual Sm–Nd and Lu–Hf fractionations which could lead them along spurious Hf–Nd isotope evolutionary paths. What we can say, however, is that the highly positive ϵ_{Nd} values ($\epsilon_{\text{Nd}} > +3$) are at the upper limit of the range of observed Nd compositions in the early Archaean. Correspondingly high radiogenic Hf compositions have not yet been recognized in these same samples. We accordingly suggest that the range of Nd compositions measured in these ancient gneisses do not faithfully represent primary isotopic variations in their source regions but rather have been produced, at least in part, by whole-rock geochemical disturbances. Until these Nd data are examined more critically, they should not be used to constrain the details of early Earth history. □

Received 7 December 1995; accepted 15 January 1996.

1. Armstrong, R. L. *Phil. Trans. R. Soc. Lond. A* **301**, 443–472 (1981).
2. Chase, C. G. & Patchett, P. J. *Earth planet. Sci. Lett.* **91**, 66–73 (1988).
3. Galer, S. J. G. & Goldstein, S. L. *Geochim. cosmochim. Acta* **55**, 227–239 (1991).
4. Bennett, V. C., Nutman, A. P. & McCulloch, M. T. *Earth planet. Sci. Lett.* **119**, 299–317 (1993).
5. McCulloch, M. T. & Bennett, V. C. *Geochim. cosmochim. Acta* **58**, 4717–4738 (1994).
6. Bowring, S. A., King, J. E., Housh, T. B., Lasachsen, C. E. & Podosek, F. A. *Nature* **340**, 222–225 (1989).
7. Collerson, K. D., Campbell, L. M., Weaver, B. L. & Palacz, Z. A. *Nature* **349**, 209–214 (1991).
8. Bowring, S. A. & Housh, T. B. *Science* **269**, 1535–1540 (1995).
9. Bridgwater, D., Rosing, M. & Schiott, L. in *Fluid Movements—Element Transport and the Composition of the Deep Crust* (ed. Bridgwater, D.) 277–298 (Kluwer, Dordrecht, 1989).
10. Rosing, M. T. *Geochim. cosmochim. Acta* **104**, 705–709 (1990).
11. Patchett, P. J. in *Proterozoic Crustal Evolution* (ed. Condie, K. C.) 481–508 (Elsevier, Amsterdam, 1992).
12. Shirey, S. B. & Hanson, G. N. *Geochim. cosmochim. Acta* **50**, 2631–2651 (1986).
13. Shirey, S. B. in *MAC Short Course on Radiogenic Isotope Systems* (eds Heaman, L. & Ludden, J. N.) 103–166 (Mineralogical Association of Canada, Nepean, 1991).
14. Jacobsen, S. B. & Dymek, R. F. *J. geophys. Res.* **93**, 338–354 (1988).
15. DePaolo, D. J. *Nature* **291**, 193–196 (1981).
16. Patchett, P. J. & Arndt, N. T. *Earth planet. Sci. Lett.* **78**, 329–338 (1986).
17. Vervoort, J. D., White, W. M. & Thorpe, R. I. *Earth planet. Sci. Lett.* **125**, 215–229 (1994).
18. Galer, S. J. G. & Goldstein, S. L. (abstr.) *Eos* **73**, 622 (1992).
19. McCulloch, M. T. & Bennett, V. C. *Lithos* **30**, 237–255 (1993).
20. Goldstein, S. L. & Galer, S. J. G. (abstr.) *Eos* **74**, 655–656 (1993).
21. Patchett, P. J. & Tatumoto, M. *Geophys. Res. Lett.* **7**, 1077–1080 (1980).
22. Patchett, P. J. *Geochim. cosmochim. Acta* **47**, 81–91 (1983).
23. Salters, V. J. M. & Hart, S. R. *Earth planet. Sci. Lett.* **104**, 364–380 (1991).
24. Fujimaki, H., Tatsumoto, M. & Aoki, K. *Proc. Lunar Planet. Sci. Conf.* **89**, B668–B672 (1984).

25. Hart, S. R. & Dunn, T. *Contr. Miner. Petrol.* **113**, 1–8 (1993).
26. Johnson, C. M. & Beard, B. L. *Nature* **362**, 441–444 (1993).
27. Gruau, G., Chauvel, C., Arndt, N. T. & Cornichet, J. *Geochim. cosmochim. Acta* **54**, 3095–3101 (1990).
28. Vervoort, J. D. & Patchett, P. J. *Geochim. cosmochim. Acta* (submitted).
29. Corfu, F. & Noble, S. R. *Geochim. cosmochim. Acta* **56**, 2081–2097 (1992).
30. Patchett, P. J., Kouvo, O., Hedge, C. E. & Tatsumoto, M. *Contr. Miner. Petrol.* **78**, 279–297 (1981).
31. White, W. M. & Patchett, P. J. *Earth planet. Sci. Lett.* **67**, 167–185 (1984).
32. Kinny, P. D., Compston, W. & Williams, I. S. *Geochim. cosmochim. Acta* **55**, 849–859 (1991).
33. Kinny, P. D. *Earth planet. Sci. Lett.* **79**, 227–247 (1986).
34. Gordon, T. M., Hunt, P. A., Bailes, A. H. & Syme, E. C. in *The Early Proterozoic Trans-Hudson Orogen of North America* (eds Lewry, J. F. & Stauffer, M. R.) 177–199 (Geological Association of Canada, St. John's, 1990).
35. Baadsgaard, H. *Earth planet. Sci. Lett.* **33**, 261–267 (1973).
36. Baadsgaard, H., Nutman, A. P. & Bridgwater, D. *Geochim. cosmochim. Acta* **50**, 2173–2183 (1986).
37. Nutman, A. P., Friend, C. R. L., Kinny, P. D. & McGregor, V. R. *Geology* **21**, 415–418 (1993).
38. Bowring, S. A., Williams, I. S. & Compston, W. *Geology* **17**, 971–975 (1989).
39. Liu, D. Y., Nutman, A. P., Compston, W., Wu, J. S. & Shen, Q. H. *Geology* **20**, 339–342 (1992).
40. Black, L. P., Williams, I. S. & Compston, W. *Contr. Miner. Petrol.* **94**, 427–437 (1986).

SUPPLEMENTARY INFORMATION. Available on Nature's World-Wide Web Site <http://www.nature.com>; requests may also be addressed to Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank H. Baadsgaard for providing us with zircons and whole-rock powders for the Greenland samples and R. Stern for the Flin-Flon sample. We thank V. Bennett for helping arrange our Hf work on the high initial ϵ_{Nd} samples. We also thank S. Galer and C. Johnson for their reviews of the manuscript. Results on the GGU samples are published with the permission of the Geological Survey of Greenland. This work was supported by a grant from the US NSF (P.J.P.).

Scales of thelodont and shark-like fishes from the Ordovician of Colorado

Ivan J. Sansom*, Moya M. Smith† & M. Paul Smith*

* School of Earth Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

† Division of Anatomy and Cell Biology, UMDS Guy's Hospital, London Bridge SE1 9RT, UK

THE phylogeny of primitive vertebrates has been investigated over the past fifteen years¹⁻³, and there has been a notable increase in the amount of data by which such phylogenies can be tested⁴⁻⁷. Stratigraphic data should provide constraints to such analyses, and we record here descriptions of putative shark and thelodont scales from the Harding Sandstone (Caradoc, Late Ordovician) of Colorado. The fossil record of sharks is extended back from the Llandovery (lower Silurian) by some 25 million years and that of thelodonts by 10 million years from the Ashgill (Late Ordovician). The new data indicate that a major radiation of lower vertebrates took place during the Ordovician.

Sharks are mainly known in the fossil record from their skeletal hard-tissue remains, generally from teeth (oral odontodes) and scales (skin odontodes)⁸. Placoid (single odontode) and polyodontoid shark scales are characterized by the presence of neck canals, emerging laterally from the pulp cavity, and polyodontoid scales are divisible into growing and non-growing types. The oldest shark remains described previously are of simple, non-growing placoid scales from the Llandovery (lower Silurian) of Siberia^{9,10}. Polyodontoid scales have also been described from the Llandovery of Mongolia¹¹ but these lack neck canals and have an unusual type of dentine, and in consequence are of uncertain taxonomic status.

Thelodonts have attracted a considerable degree of attention in recent years, with the description of a number of new forms from the Silurian of Canada suggesting the presence of fairly advanced characteristics in at least one lineage⁵. Previously the earliest recorded thelodonts were from the Ashgill (Late Ordovician) of Siberia¹².

The material we describe here is from the Harding Sandstone of

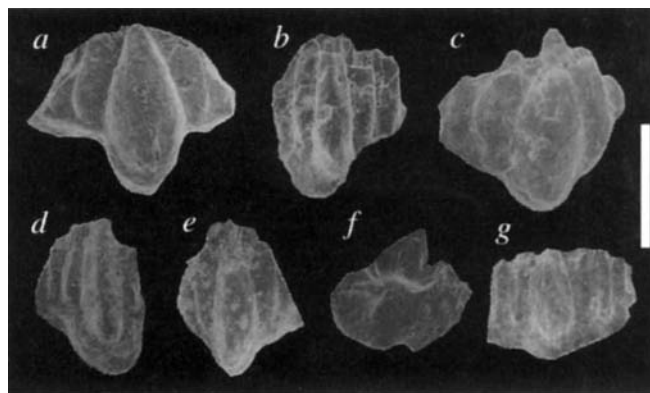
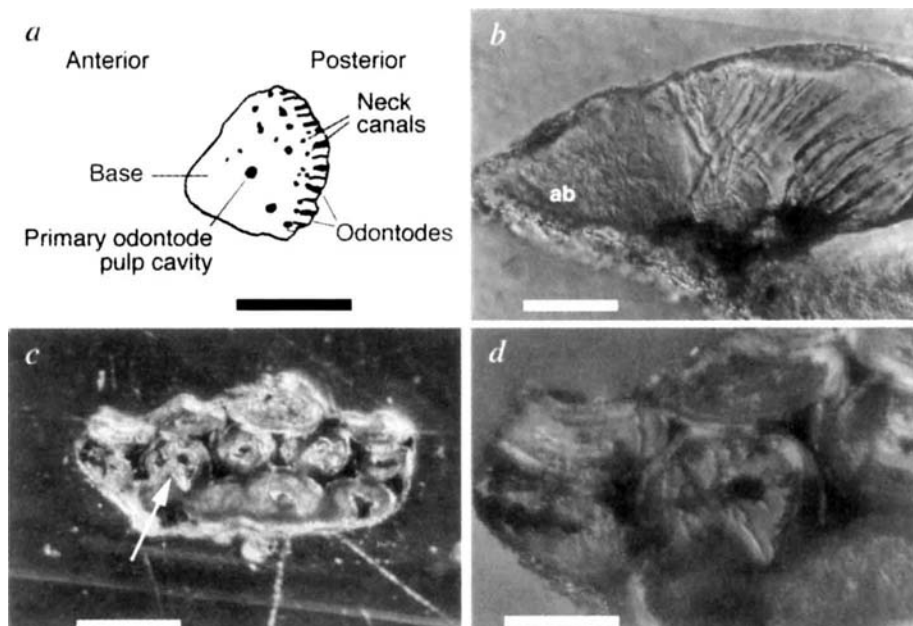


FIG. 1 Representatives of scale morphology A, showing the presence of polyodontodes and the tear-drop shaped primordial odontode (especially prominent in a and c). All crown views of scales apart from f which is basal. Scale bar, 500 μ m. a, specimen BU 2574; b, BU 2575; c, BU 2576; d, BU 2577; e, BU 2578; f, BU 2579; g, BU 2580. Specimens are deposited in the Lapworth Museum, School of Earth Sciences, University of Birmingham, UK.

Colorado, which crops out around Cañon City, some 50 km southwest of Colorado Springs. The unit has been known to yield vertebrate remains for over 100 years¹³, and dates from the basal Franklinian Stage of the Mohawkian Series, equivalent to the early middle Caradoc of Britain. Previous studies of fish from the Harding Sandstone have identified the presence of two common forms, *Astraspis* and *Eriptychius*, which are thought to be distantly related to the Silurian heterostracans. A third form, which until recently was only known from its distinct cellular histology in thin sections of the Harding Sandstone bone beds¹⁴ may represent a primitive acanthodian (M. M. S. and I. J. S., manuscript submitted). The presence of additional taxa in the formation has been alluded to previously¹⁵, but these have never been documented and are rarely cited.

Scale morphology A from the Harding conforms to features exhibited by chondrichthyan placoid scales in every respect. The scales are formed from a crown consisting of multiple odontodes which show a cyclomorial growth pattern, with a large, tear-drop shaped primordial odontode (Fig. 1). Growth is evidenced by the

FIG. 2 Histology of scale morphology A. a, Camera lucida drawing of the basal view of BU 2581 showing pulp and neck canal openings. The number of pulp openings always exceeds the number of odontodes in scale morphology A. Scale bar, 500 μ m. b, Longitudinal double-polished thin section (BU 2582) through primary odontode, showing acellular base (ab) overlain by a crown of coarse tubular dentine with a single pulp cavity. Scale bar, 100 μ m. c, Transverse double-polished section (BU 2583) showing stacked polyodontodes towards the posterior of the scale. Arrow marks odontode shown in d with a fine pulp projection, presumable a neck canal. Scale bars: c, 100 μ m; d, 50 μ m.



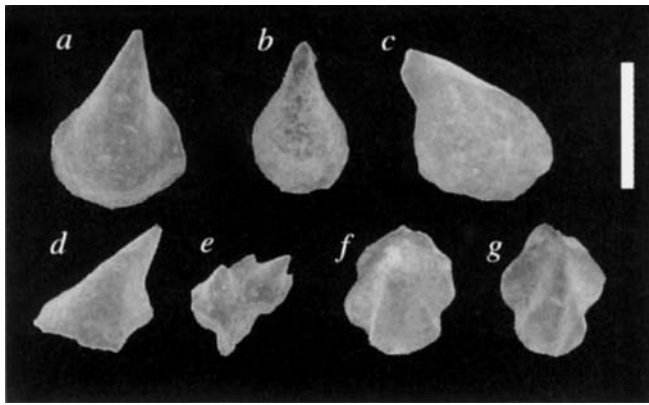


FIG. 3 a, b, Representatives of scale morphology B: conical specimens with a very thin base. Crown views of specimens BU 2584 (a) and BU 2585 (b). c, Basal view of BU 2586, d, Antero-lateral view of BU 2587. e, Denticulated plate (BU 2588) found in association with scale morphologies B and C. f, g, Crown views of scale morphology C (BU 2589 and BU 2590, respectively). Scale bar, 500 μ m.

position of overlapping smaller odontodes of similar morphology, lateral and posterior to the primary site of odontode morphogenesis (Fig. 2c). These scales correspond to 'Type C' chondrichthyan scales¹⁰, and are characterized histologically by an arboreal branching form of tubular dentine (Fig. 2b), with slender neck canals emerging from the posterior of the elongate pulp cavity (Fig. 2a, d).

The cyclomorall growth exhibited in scale morphology A differs notably from the simple, single placoid scales from which *Elegestolepis* (Llandovery), otherwise the oldest putative chondrichthyan, is known^{9,10}. Scale morphology A also lacks a glassy cap of enameloid or vitrodentine which is present in *Elegestolepis*. Mongolepid scales, from the Llandovery of Mongolia and Siberia¹¹, have crowns formed by polyodontodes but these represent synchronous growth and, histologically, comprise an atubular form of dentine termed lamellin. Mongolepid scales also differ from typical chondrichthyan scales in lacking neck canals and instead have a horizontal canal system between individual odontodes. In the absence of further characters, notably of the body plan and tooth morphology, it seems unlikely that a firm phylogenetic

relationship could be proposed between these Lower Palaeozoic taxa all of which lack associated teeth and positive evidence for the presence of jaws¹⁶, and the major groups of Upper Palaeozoic sharks.

We have also been able to identify scales with morphological and histological characteristics that identify them as loganiid thelodonts. The most abundant (scale morphology B) are conical in shape, with the point of the cone lying at a very shallow angle to the rounded and flared base (Fig. 3a–d). A broad, irregular pulp cavity opens onto the visceral surface and a narrow posterior pulp canal runs along the internal part of the cone (Fig. 4a). Scale morphology C is represented by squat rhomboid scales with serrated margins produced by a series of prominent, short ridges (Fig. 3f–g). The scale crown is formed from fine, branching tubular dentine, also characteristic of scale morphology B (Fig. 4b, d), but instead of a single pulp cavity the basal half of scale morphology C has numerous wide dentine canals from which the finer tubules emerge (Fig. 4c). This type of dentine is characteristic of loganiid thelodonts¹⁷. Extremely small denticulated plates are found in association with the thelodont scales (Fig. 3e); their size and the presence of slender cusped denticles is suggestive of the denticle-bearing plates described from inside the snout, pharyngeal and branchial opening regions of articulated loganiid thelodonts from the Llandovery of Scotland^{18,19}. Scale morphologies B and C follow the characteristics of body and head scales, respectively, in other loganiids.

Recent work has postulated that achanolepid-type thelodonts, which possess scales formed from a fine tubular dentine and lacking pulp canals, represent the primitive condition for thelodonts¹⁷. The presence of broad dentine canals in the base of the head scales of the Harding Sandstone thelodont and the short posterior pulp canal in the body scales both indicate that loganiid thelodonts first appeared before the achanolepids.

In addition to the putative chondrichthyan and thelodont scales described above, at least six other undescribed taxa are present within our new collections from the Harding Sandstone although their classification is, at present, enigmatic²⁰. The presence of sharks, thelodonts, possible acanthodians, conodonts and heterostracan-like fish in the Caradoc indicates that the major period of diversification within the lower vertebrates was already well underway during the Ordovician, not the Silurian as has been suggested classically, and that fish evolution was not slow during the Ordovician^{21,22}. □

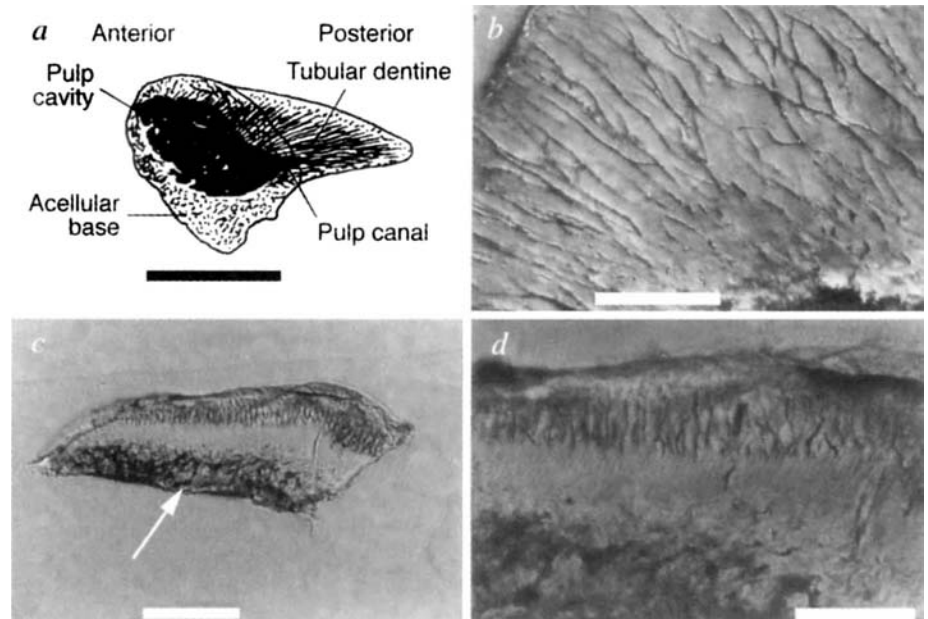


FIG. 4 a, b, Histology of scale morphology B. a, Camera lucida drawing of a three-dimensional specimen (BU 2591) viewed through transmitted light whilst immersed in clove oil. Specimen shows thin, acellular base, broad pulp cavity, posterior pulp canal and fine tubular dentine. Scale bar, 500 μ m. b, Longitudinal double-polished thin section (BU 2592) showing branching fine tubular dentine. c, d, Histology of scale morphology C. c, Longitudinal double-polished thin section (BU 2593) with coarse dentine canals in the base (arrowed) overlaid by fine tubular dentine. Scale bar, 100 μ m. d, Detail of c showing fine tubular dentine. Scale bar, 50 μ m.

Received 20 October; accepted 11 December 1995.

1. Janvier, P. J. *vert. Paleont.* **1**, 121–159 (1981).
2. Forey, P. & Janvier, P. *Nature* **361**, 129–134 (1993).
3. Blicek, A. *Geobios* **25**, 101–113 (1992).
4. Gagnier, P.-Y., Blicek, A. & Rodrigo, S. G. *Geobios* **19**, 629–634 (1986).
5. Wilson, M. V. H. & Caldwell, M. W. *Nature* **361**, 442–444 (1993).
6. Sansom, I. J., Smith, M. P., Armstrong, H. A. & Smith, M. M. *Science* **256**, 1308–1311 (1992).
7. Aldridge, R. J., Briggs, D. E. G., Smith, M. P., Clarkson, E. N. K. & Clark, N. D. L. *Phil. Trans. R. Soc. Lond. B* **338**, 405–421 (1993).
8. Zangerl, R. in *Handbook of Paleichthyology* 3 (ed. Schultz, H.-P.) 1–113 (Gustav Fischer, Stuttgart, 1981).
9. Karatajuté-Talimaa, V. *Palaeontographica A* **143**, 35–50 (1973).
10. Karatajuté-Talimaa, V. in *Fossil Fishes as Living Animals* (ed. Mark-Kurik, E.) 223–231 (Academy of Sciences of Estonia, 1992).
11. Karatajuté-Talimaa, V., Novitskaya, L. I., Rozman, H. S. & Sodov, Z. *Paleont. J.* **1**, 76–86 (1990).
12. Karatajuté-Talimaa, V. *Modern Geol.* (in the press).
13. Walcott, C. D. *Bull. geol. Soc. Am.* **3**, 153–171 (1892).
14. Denison, R. H. *Fieldiana, Geol.* **16**, 131–192 (1967).
15. Spjeldnaes, N. *Colloques int. Cent. natn. Rech. scient.* **163**, 11–20 (1967).
16. Long, J. A. *The Rise of Fishes* (Johns Hopkins, Baltimore, 1994).
17. Turner, S. in *Early Vertebrates and Related Problems of Evolutionary Biology* (eds Chang, M.-M. & Zhang, G.-R.) 87–119 (Science, Beijing, 1991).
18. Van der Bruggen, W. & Janvier, P. *Nature* **364**, 107 (1993).
19. Vergoosen, J. M. J. *Geologie en Mijnbouw* **71**, 51–64 (1992).
20. Sansom, I. J., Smith, M. M., Smith, M. P. & Turner, P. *Geobios* **MS 19**, 57–59 (1995).
21. Benton, M. J. *Vertebrate Palaeontology* (Harper Collins, London, 1990).
22. Carroll, R. L. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1987).

ACKNOWLEDGEMENTS. This work was supported by a NERC research grant. We thank P. Turner, J. Keeton, R. Watters and C. Fair for their assistance while in the field; V. Karatajuté-Talimaa, S. Turner and J. Vergoosen for their comments on the material; and P. Forey and P. Janvier for their comments on the manuscript.

Increased soldier production in ant colonies exposed to intraspecific competition

Luc Passera*, Eric Roncin*, Bernard Kaufmann* & Laurent Keller†‡

* Laboratoire d'Ethologie et Psychologie animale, URA CNRS 1837 Université Paul-Sabatier, 118 route de Narbonne, 31060 Toulouse Cedex, France

† Institut de Zoologie et d'Ecologie Animale, University of Lausanne, BB, 1015 Lausanne and Zoologisches Institut, Bern University, Ethologische Station Hasli, 3032 Hinterkappelen, Switzerland

THE success of organisms and their increasing complexity throughout the course of evolution are thought to have depended on a small number of important transitions, one of which was the shift from a solitary lifestyle to societies of organisms exhibiting division of labour and having specialized castes^{1,2}. The most familiar examples of the advantages arising from division of labour and caste differentiation come from social insects³. It has been suggested that the proportion of workers of various physical castes has evolved to enhance the fitness of colony members^{4,6} with the prediction that caste ratios should vary with environmental factors such as predation, competition and food availability. However, because it has been difficult to manipulate experimentally the environmental factors believed to influence optimal caste ratio, demonstration of adaptive caste distribution has remained elusive^{7–10}. Here we show that colonies of the ant *Pheidole pallidula* increase the relative investment in soldier production after perceiving the presence of foreign conspecific colonies. This is the first experimental demonstration of a social insect altering physical caste ratios in an adaptive manner.

The worker force of *Pheidole pallidula* is divided into two discrete morphological castes¹¹: small-headed minor workers, who carry out most of the work (such as foraging and brood care), and large-headed major workers (soldiers), specialized for defence¹². In natural colonies, soldiers make up between 3.4% and 41% of the worker force (mean $\pm$ s.d. = $9.8 \pm 4.5\%$, $n = 72$).

Because soldiers seldom perform duties other than colony defence, the relative proportion of soldiers should increase when competition generated by the presence of nearby conspecific colonies^{13,14} is more likely⁴.

To test whether colonies respond to the presence of conspecific colonies by increasing soldier production, we set up two groups of colonies. In the first group (test colonies) workers could perceive the presence of workers of a foreign colony in a tunnel (85 mm length, 4 mm wide) that was divided over its whole length by a fine wire mesh. Workers perceived the presence of foreign-colony workers as they passed through the tunnel to the foraging arena, where food was provided. This experimental set-up mimicked a natural situation where interspecific interactions occur on foraging trails. Workers could only insert their antennae and legs through the mesh, so mortal injuries could not occur. Previous studies have shown that ant workers perceive and react to the presence of foreign conspecifics under such experimental conditions¹⁵. For the control treatment, the tunnel was partitioned by a

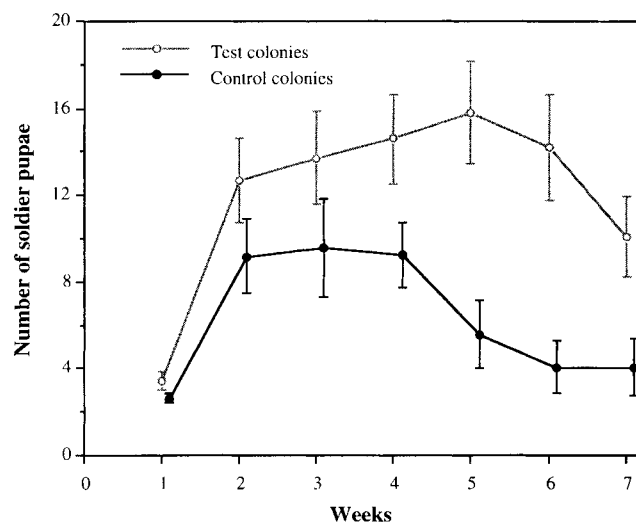


FIG. 1 Mean number ($\pm$ s.e.m.) of soldier pupae in colonies where workers had contact with foreign workers (test colonies, $n = 17$) or no contact with foreign workers (control colonies, $n = 16$) during the 7 weeks following the production of the first soldier pupae. A repeated-measures ANOVA (all data were square-root transformed to give normal distributions¹⁶) showed that the number of soldier pupae was higher in test colonies than control colonies ($F_{1,31} = 20.04$, $P < 0.001$) with a significant difference in the number of soldier pupae over time ($F_{6,186} = 10.06$, $P < 0.001$). The significant interaction between treatment and time ($F_{6,186} = 2.20$, $P < 0.05$) reflects an increasing difference over time between test and control colonies in the difference of the number of soldier pupae. We collected 40 large colonies of *Pheidole pallidula* with their queen near Montauban in southern France in spring 1994. Of these colonies, 20 were randomly assigned to the test treatment and 20 to the control treatment. The initial proportion of soldiers was similar in test (10.3%) and control colonies (12.4%; $t = 0.72$, $P = 0.48$). All soldiers were removed from these colonies and the number of minor workers standardized to 1,100 (± 100). Colonies were maintained under standard conditions at 27 °C (ref. 27) and fed *ad libitum* on honey and mealworms in the foraging box. The workers of test colonies could perceive the presence of workers from another test colony through the wire mesh when travelling to the foraging arena. The frequency of contacts between foreign workers remained stable throughout the experiment, occurring on about 40% of each of the worker foraging trips. Three test and four control colonies were subsequently excluded from the analysis because the queen died during the experiment. Colonies were censused weekly during the 7 weeks after the eclosion of the first soldier pupae. The first soldier pupae in test colonies were produced 48 ± 9 ($n = 17$) days after colonies were set up, not significantly different from control colonies (45 ± 3 days; $n = 16$; $t = 1.18$, NS). The decrease in the number of soldier pupae after week five stems from the inhibitory effect that the presence of adult soldiers has on the production of the new soldiers¹¹.

‡ To whom correspondence should be addressed.

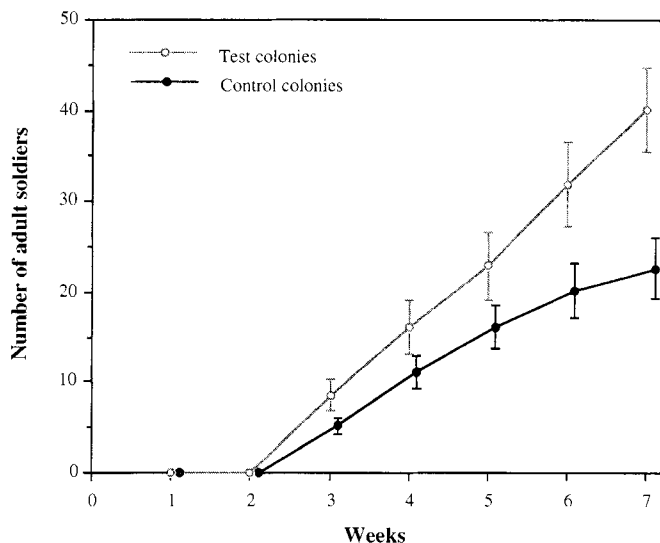


FIG. 2 Mean number ($\pm$ s.e.m.) of adult soldiers in colonies where workers had contact (test colonies) or no contact with foreign workers (control colonies) during the 7 weeks after the production of the first soldier pupae. The number of soldiers was higher in test colonies than control colonies ($F_{1,31} = 5.05$, $P < 0.05$) with a significant difference in the number of soldiers over time ($F_{6,186} = 95.75$, $P < 0.001$). The significant interaction between treatment and time ($F_{6,186} = 2.26$, $P < 0.05$) reflects the increasing difference over time between test and control colonies in the number of adult soldiers. Methods, statistics and sample sizes as in Fig. 1

plastic sheet 1 mm thick, preventing contact between workers of foreign colonies.

The production of soldier pupae increased at a faster rate and remained significantly higher in test colonies compared with control colonies (Fig. 1). Overall, the total number of soldier pupae censused over the 7 weeks was nearly twice as high in test colonies (mean $\pm$ s.d. = 84.8 ± 29.6 , $n = 17$) as in control colonies (43.6 ± 24.1 , $n = 16$; t -test on square-root transformed values¹⁶: $t = 4.39$, $P < 0.001$). The higher production of soldier pupae in test colonies resulted in a significantly greater increase in the number of adult soldiers (Fig. 2). At the end of the experiment the number of adult soldiers in test colonies (40.1 ± 19.2 , $n = 17$) was nearly twice as high as in controls (22.6 ± 13.2 ; $n = 16$; t -test on square-root transformed values: $t = 3.03$, $P = 0.002$). But the two types of colonies did not differ in total biomass (dry weight of adult minor workers + soldiers) at the end of the experiment (test colonies, 109.4 ± 29.1 mg, $n = 17$; control colonies, 105.6 ± 20.0 mg, $n = 16$; t -test on square-root transformed values: $t = 0.32$, $P = 0.75$).

The outcome of these differences in resource allocation was a significantly higher relative biomass of adult soldiers in test colonies ($27.6 \pm 11.2\%$; $n = 17$) than in control colonies ($15.7 \pm 9.6\%$; $n = 16$; t -test on square-root arcsine transformed proportions¹⁸: $t = 3.06$, $P = 0.004$) at the end of the experiment. The greater relative investment by test colonies in soldier production was also reflected in the higher relative biomass of soldier pupae (measured as the ratio of dry weight of soldier pupae to dry weight of all pupae produced during the 7 weeks of the experiment, see ref. 17) in test colonies ($21.7 \pm 7.5\%$; $n = 17$) compared with control colonies ($13.2 \pm 7.7\%$; $n = 16$; t -test on square-root arcsine transformed proportions: $t = 3.20$, $P = 0.003$). Together these data indicate that colonies reacted to the presence of conspecific colonies by allocating a greater proportion of energy to soldier production, resulting in a greater number of soldiers.

Morphological differences between minor workers and soldiers arise from a different developmental program during the larval stage¹⁸. For example, soldier induction occurs during the two last larval stages in *Pheidole bicarinata*¹⁹, the underlying mechanism

being a nutritional switch which is translated into endocrine signals that coordinate subsequent patterns of differentiation¹⁸. The switch between the soldier and minor worker developmental pathways is also influenced by trophic factors in *P. pallidula*¹¹. The higher proportion of worker larvae that followed the soldier developmental pathway thus probably arose from a shift in the behaviour of nurse workers feeding larvae after the presence of the foreign colony was perceived. The ability of workers to distinguish members of their own colony from foreign individuals is well known in ants²⁰. Discrimination is generally based on colony-specific chemical cues that are located on the cuticle of the individuals²⁰. Therefore, the most likely mechanisms by which test colonies perceived the presence of foreign colonies is by workers touching the bodies of foreign workers with their antennae.

Our results show that, as predicted by caste ratio theory, *P. pallidula* colonies increase soldier production in response to the presence of foreign workers. This is the first experimental evidence of a social insect altering physical caste ratios in relation to environmental factors. There are two possible explanations as to why the variation in environmental factors did not elicit an adaptive response in previous studies⁷⁻¹⁰. First, flexible caste ratio is expected to evolve only when the fluctuation in environmental factors such as competition is longer than the average interval between the moment of caste determination and adult eclosion⁴, which possibly was not the case for species previously studied. Alternatively, colony response might be elicited by cues other than those that were quantified or manipulated in these experiments.

The increased production of soldiers in response to the presence of foreign workers provides a striking case of adaptive phenotypic plasticity²¹⁻²⁵ which is shaped by natural selection acting simultaneously at the individual and colony level^{2,4,26}. Individual- and colony-level selection may have opposing effects when differences in caste membership correlate with differences in reproductive opportunities. But because both *P. pallidula* soldiers and minor workers lack ovaries²⁷, and hence entirely surrender their personal reproductive potential, they can ensure the transmission of copies of their genes only by contributing to colony success²⁸. Therefore, there is no conflict between individual- and colony-level selection regarding the optimal ratio of these two castes^{4,25,29}, a condition which may have facilitated the evolution of flexible caste ratios. □

Received 25 September; accepted 15 December 1995.

- Smith, A. *The Wealth of Nations* (Penguin, London, 1776).
- Szathmari, E. & Maynard Smith, J. *Nature* **374**, 227–231 (1995).
- Wilson, E. O. *Science* **190**, 798–800 (1975).
- Oster, G. & Wilson, E. O. *Caste Ecology in the Social Insects* (Princeton Univ. Press, 1978).
- Wilson, E. O. *Behav. Ecol. Sociobiol.* **16**, 89–98 (1984).
- Wilson, E. O. *Science* **228**, 1489–1495 (1985).
- Johnston, A. B. & Wilson, E. O. *Ann. ent. Soc. Am.* **78**, 8–11 (1985).
- Calabi, P. & Traniello, J. F. A. *Behav. Ecol. Sociobiol.* **24**, 69–78 (1989).
- Hasegawa, E. *Behav. Ecol. Sociobiol.* **33**, 73–77 (1993).
- Beshers, S. N. & Traniello, J. F. A. *Ecology* **75**, 763–775 (1994).
- Passera, L. & Suzzoni, J. P. in *Morphogenetic Hormones of Arthropods* (ed. Gupta, A. P.) 400–430 (Rutgers Univ. Press, New Brunswick, 1991).
- Detrain, C. & Pasteels, J. M. *Behav. Ecol. Sociobiol.* **29**, 405–412 (1992).
- Hölldobler, B. & Lumsden, C. J. *Science* **210**, 732–739 (1980).
- Gordon, D. M. *Behav. Ecol. Sociobiol.* **31**, 417–427 (1992).
- Nonacs, P. & Calabi, P. *Proc. R. Soc. Lond. B* **249**, 95–99 (1992).
- Zar, H. J. *Biostatistical Analysis* (Prentice-Hall, Englewood Cliffs, 1984).
- Aron, S., Passera, L. & Keller, L. *J. evol. Biol.* **7**, 403–418 (1994).
- Wheeler, D. E. *Am. Nat.* **138**, 1218–1238 (1991).
- Wheeler, D. E. & Nijhout, H. F. *Science* **213**, 361–363 (1981).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Springer, Berlin, 1990).
- West-Eberhard, M. J. A. *Rev. Ecol. Syst.* **20**, 249–278 (1989).
- Keller, L. & Ross, K. G. *Science* **260**, 1107–1110 (1993).
- Keller, L. & Ross, K. G. *Behav. Ecol. Sociobiol.* **33**, 121–129 (1993).
- Pfennig, D. W. & Collins, J. P. *Nature* **362**, 836–838 (1993).
- Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
- Ratnieks, F. L. W. & Reeve, H. J. *theor. Biol.* **158**, 33–65 (1992).
- Passera, L. *Insectes soc.* **25**, 117–126 (1978).
- Hamilton, W. D. J. *theor. Biol.* **7**, 1–52 (1964).
- Maynard Smith, J. in *Evolutionary Progress* (ed. Nitecki, M. H.) 219–223 (Univ. Press of Chicago, 1988).

ACKNOWLEDGEMENTS. We thank G. Bernasconi, T. Bakker, K. Boomsma, A. Bourke, P. Christe, N. Franks, B. Hölldobler, M. Milinski, P. Nonacs, F. Ratnieks, K. Ross, L. Sundström and E. Vargo for comments on the manuscript. This work was funded by the Ministère français de l'Enseignement supérieur et de la Recherche (E.R.) and grants from the Swiss National Science Foundation (L.K.)

Abnormal splicing of the leptin receptor in *diabetic* mice

Gwo-Hwa Lee^{*†}, R. Proenca^{*†}, J. M. Montez[†],
K. M. Carroll[†], J. G. Darvishzadeh[†], J. I. Lee^{*}
& J. M. Friedman^{*†}

^{*} Howard Hughes Medical Institute, and [†] Department of Molecular Genetics, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

MUTATIONS in the mouse *diabetes* (*db*) gene result in obesity and diabetes in a syndrome resembling morbid human obesity¹. Previous data suggest that the *db* gene encodes the receptor for the *obese* (*ob*) gene product, leptin²⁻⁷. A leptin receptor was recently cloned from choroid plexus and shown to map to the same 6-cM interval on mouse chromosome 4 as *db*⁸. This receptor maps to the same 300-kilobase interval as *db*, and has at least six alternatively spliced forms. One of these splice variants is expressed at a high level in the hypothalamus, and is abnormally spliced in C57BL/Ks *db/db* mice. The mutant protein is missing the cytoplasmic region, and is likely to be defective in signal transduction. This suggests that the weight-reducing effects of leptin may be mediated by signal transduction through a leptin receptor in the hypothalamus.

A physical map of the *db* gene was compiled using the microdissection clone D4Rck22 as the starting point of a chromosome walk⁹. The mutation was localized to a 300-kb interval covered by two bacterial artificial chromosomes (BACs), 43 and 242 (Fig. 1a). Two previously cloned human genes that flank *db*, *JAK1* and

PDE48, map to human chromosome 1p31, suggesting that the human homologue of the *db* gene maps to this chromosomal region^{10,11}.

Candidate genes for *db* were isolated from BACs 43 and 242 by using exon trapping, and complementary DNA selection from mouse hypothalamus^{12,13}. A mouse-brain cDNA library was screened with seven putative gene fragments. Each of nine cDNA clones contained sequences identical to the leptin receptor, Ob-R⁸. The position of Ob-R in the contig was determined using polymerase chain reaction (PCR) assays specific for the 5' and 3' end of the cDNA (Fig. 1b). The gene spans ~ 100 kb and is transcribed towards the telomere.

Several cDNA clones diverge from Ob-R at the carboxy terminus, and two diverge at the extreme amino terminus. In four cases, the predicted C-terminal sequences were identical until Lys 889 of the leptin receptor which includes the transmembrane domain. The numbers of the amino acids refer to the published Ob-R sequence⁸. Beyond this point, the cDNAs predict proteins with differing cytoplasmic portions, designated Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rd (Fig. 2). Ob-Re predicted a different amino-acid sequence after His 796, and may encode a soluble receptor (Fig. 2). Ob-Ra has been shown to correspond to mouse Ob-R⁸. The C terminus of Ob-Rb was 78% identical to the human receptor, suggesting that it is the mouse homologue⁸. The Ob-Rb cDNA a sequence predicts a potential 'Box 2' sequence (underlined in Fig. 2), a motif required for the binding of JAK protein kinases¹⁴⁻¹⁶. Ob-Rb is also alternatively spliced at the N terminus. Although one form of Ob-Rb has the extracellular region of Ob-R, another clone diverged upstream of Pro 664. Identical N-terminal sequences were also seen in a splice variant of Ob-Rc. However, in both cases the divergent N-terminal sequences predicted stop codons in all three reading frames.

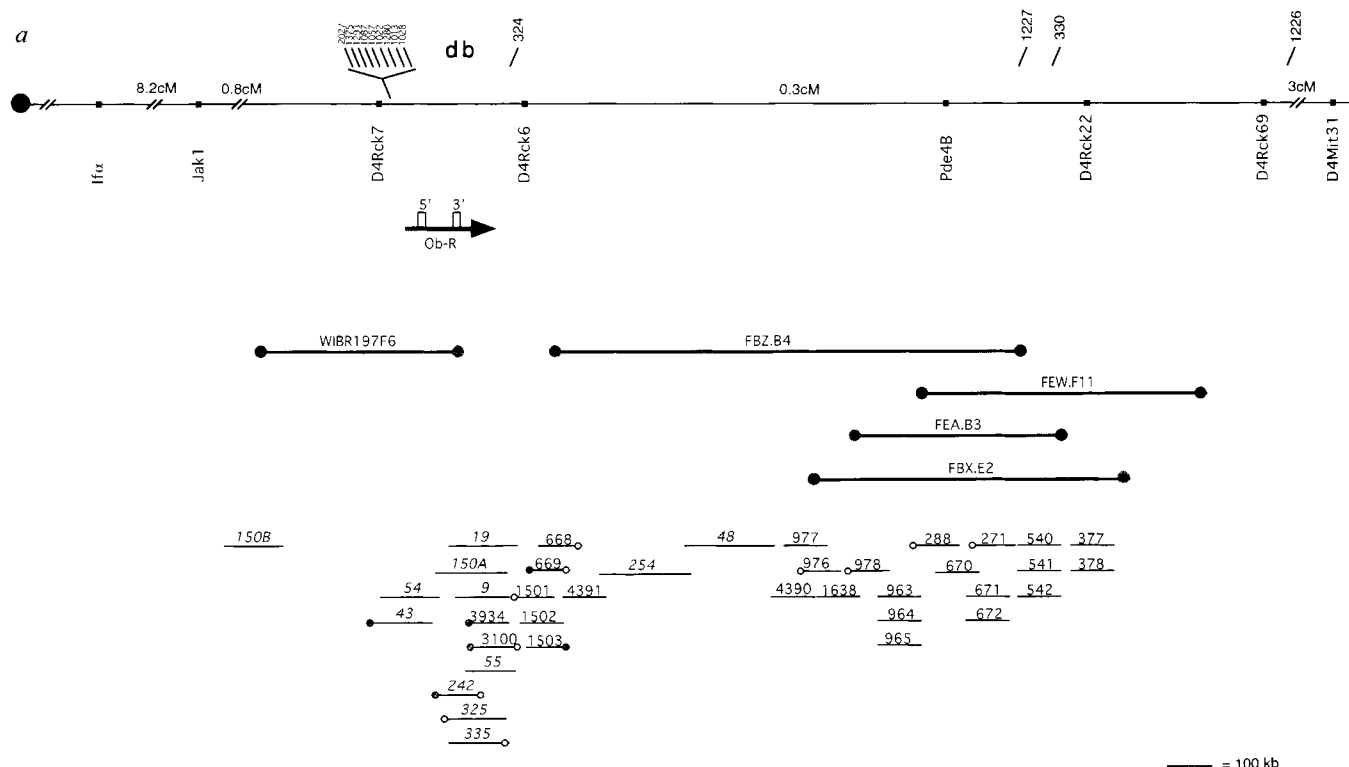


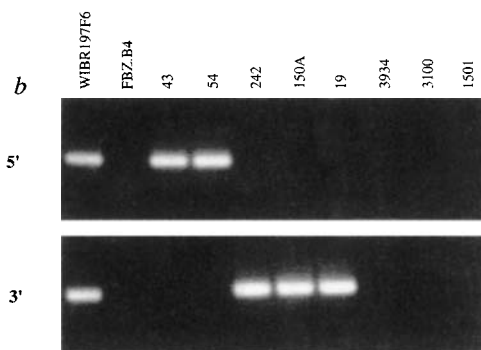
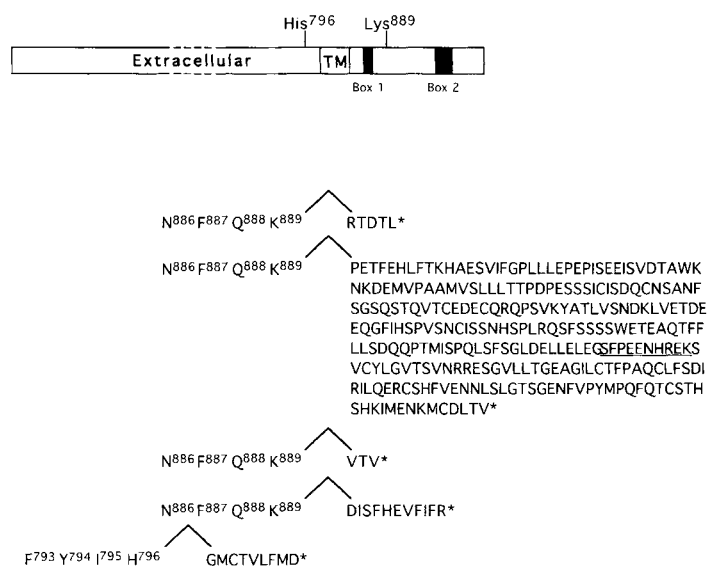
FIG. 1 Localization of the leptin receptor to the region of the *db* gene. *a*, The *db* mutation was segregated in two crosses, a C57BL/Ks *db/db* × *Mus spretus* intercross and a C57BL/Ks *db/db* × *Mus castaneus* intercross

(unpublished data) totalling 1150 meioses²⁰. A genetic map was compiled ► by genotyping the progeny of these crosses with the markers indicated (top). Recombinant mice are noted on the map as numbers. A chromosome walk

FIG. 2 The leptin receptor is alternatively spliced.

A total of nine leptin receptor cDNAs were isolated from mouse brain, and were found to encode several splice variants. A representation of the leptin receptor is shown at the top, including putative motifs for the transmembrane domain (TM) and those for JAK binding, 'Box 1' and 'Box 2' (refs 15,16). The numbers of the amino acids refer to the published Ob-R sequence⁸. Six of the clones had identical sequences until Lys 889, at which point the predicted proteins diverged. These mRNAs are designated Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rd, and their predicted amino-acid sequences are shown. Ob-Ra corresponds to the published mouse Ob-R, and Ob-Rb corresponds to the human OB-R⁸. Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rd predict a Box 1 motif, and Ob-Rb also predicts a peptide sequence potentially homologous to Box 2 (underlined). Two independent cDNA clones were identical to the leptin receptor upstream of His 796, at which point the nucleotide sequences diverged (Ob-Re). Its nucleotide sequence predicts a soluble receptor.

METHODS. cDNA selection was performed using mouse hypothalamus RNA¹³. Library screening, exon trapping and DNA sequencing were performed as described previously³. The C-terminal sequences of Ob-Ra, Ob-Rc, Ob-Rd and Ob-Re were found in cDNA clones. The C-terminal sequence of Ob-Rb was not full length. The C-terminus sequence was initially completed by sequencing genomic DNA. The template was prepared using vectorette PCR of BAC 242 with primers from the cDNA²³. The sequence was confirmed by sequencing RT-PCR products.



was initiated with the microdissection clone D4Rck22. The walk spanned 2.7 Mb and was composed of yeast artificial chromosomes (YACs, bold lines), bacterial artificial chromosomes (BACs, italics) and P1 bacteriophage (numbers). Genotyping of the recombinant animals with two single strand length polymorphism (SSLP) markers, D4Rck6 and D4Rck7, localized the *db* gene to the ~300-kb interval between the recombination events in mice 324 and 1028; this interval was spanned by BACs 242 and 43. *b*, PCR indicated that the 5' end of the leptin receptor mapped to BACs 43 and 54. The 3' end of Ob-R mapped to BACs 19, 150A and 242, indicating that the gene is transcribed towards the telomere (see *a*).

METHODS. YAC clones and YAC ends were isolated as described previously^{3,21-24}. P1 clones were isolated by sending specific pairs of PCR primers to Genome Systems (St Louis, MO)²⁵. P1 ends were recovered using vectorette PCR²⁶. BACs were isolated as described previously²⁷. Primer selection and PCR amplification were performed as described previously³. The primers were: D4Rck6f 5'-ATCTTGGGTCTCTGAAGAA-3', D4Rck6r 5'-GAGATTGTCAGTCACAGCCTC-3', D4Rck7f 5'-ATCTGAATTGGAATCAATACAC-3', D4Rck7r 5'-AAATCTGTATCCTTCTGAAAC-3'. The primers for the 5' end of Ob-R were f5'-TTCAGAAGCCCCCTTCAAAG-3' and r5'-GAATTCCTTATGTGATAGCTGC-3'. The 3' end primers were f5'-TGGATTAACCCCTGCTCTCA-3' and r5'-GGTCTCAGAGCACCCAGGTA-3'.

C57BL/Ks *db/db* mice have a longer fragment length of Ob-Rb C-terminal reverse transcription (RT)-PCR products from hypothalamic RNA (Fig. 3a). However, PCR-amplified genomic DNA spanning the splice acceptor at Pro 890 was of normal size and wild-type sequence in C57BL/Ks *db/db* mice (Fig. 3a). In addition, both the size and nucleotide sequence of RT-PCR products corresponding to the other 3' ends were normal in the *db* mice, suggesting that the splice donor at Lys 889 is also normal (Fig. 3b). These data suggest that the longer Ob-Rb fragment from C57BL/Ks *db/db* mice was the result of abnormal splicing.

Sequencing of the RT-PCR products of Ob-Rb from the mutant mice indicated a 106-bp insertion between the splice donor at Lys 889 and the splice acceptor at Pro 890. The sequence of the inserted DNA was identical to the first 106 bp of the C-terminal Ob-Ra exon downstream of its splice acceptor at Arg 890 (Fig. 3c). Sequencing of genomic DNA and RT-PCR products from the 3' untranslated region of Ob-Ra from C57BL/Ks *db/db* mice identified a G → T mutation 108 bp after the splice acceptor (Fig. 3d), which results in the appearance of a consensus splice donor site, AGGTAAA¹⁷. This mutant splice donor results in the splicing of 106 bp of the Ob-Ra terminal exon into the splice acceptor at Pro 890 of Ob-Rb RNA. The mutant Ob-Rb protein has a termination codon five amino acids after the splice, and is missing most of the cytoplasmic region, including the potential Box 2 motif^{15,16}.

The Ob-Rb leptin receptor is expressed at a high level in the hypothalamus relative to other tissues (Fig. 4). Lower-level expression is seen in brain and testes, with an even lower level in adipose tissue. The other alternatively spliced messenger RNAs are expressed in several tissues including, in some cases, hypothalamus (Fig. 4). Ob-Re, which may encode a soluble receptor, is expressed in adipose tissue, hypothalamus, heart and testes.

The C57BL/Ks *db/db* mutation is co-isogenic and results in the functional replacement of Ob-Rb by Ob-Ra. These data, combined with the mapping of the leptin receptor to precisely the same chromosomal region as *db*, strongly suggest that the cytoplasmic region of Ob-Rb is allelic with *db*. Although the Ob-Rb C terminus is altered in *db* mice, it is possible that the mutant splice donor of Ob-Ra is also inserted into other (not yet identified) transcripts. It is also possible that the peptide sequence in the extracellular region of the mutant protein diverges from that of Ob-R as a result of alternative splicing. The identification of mutations in the two other available alleles of *db* (*db^{pas}* and *db³¹*) may provide additional information on the peptide sequence of the mutant protein, as well as its structure-function relationship².

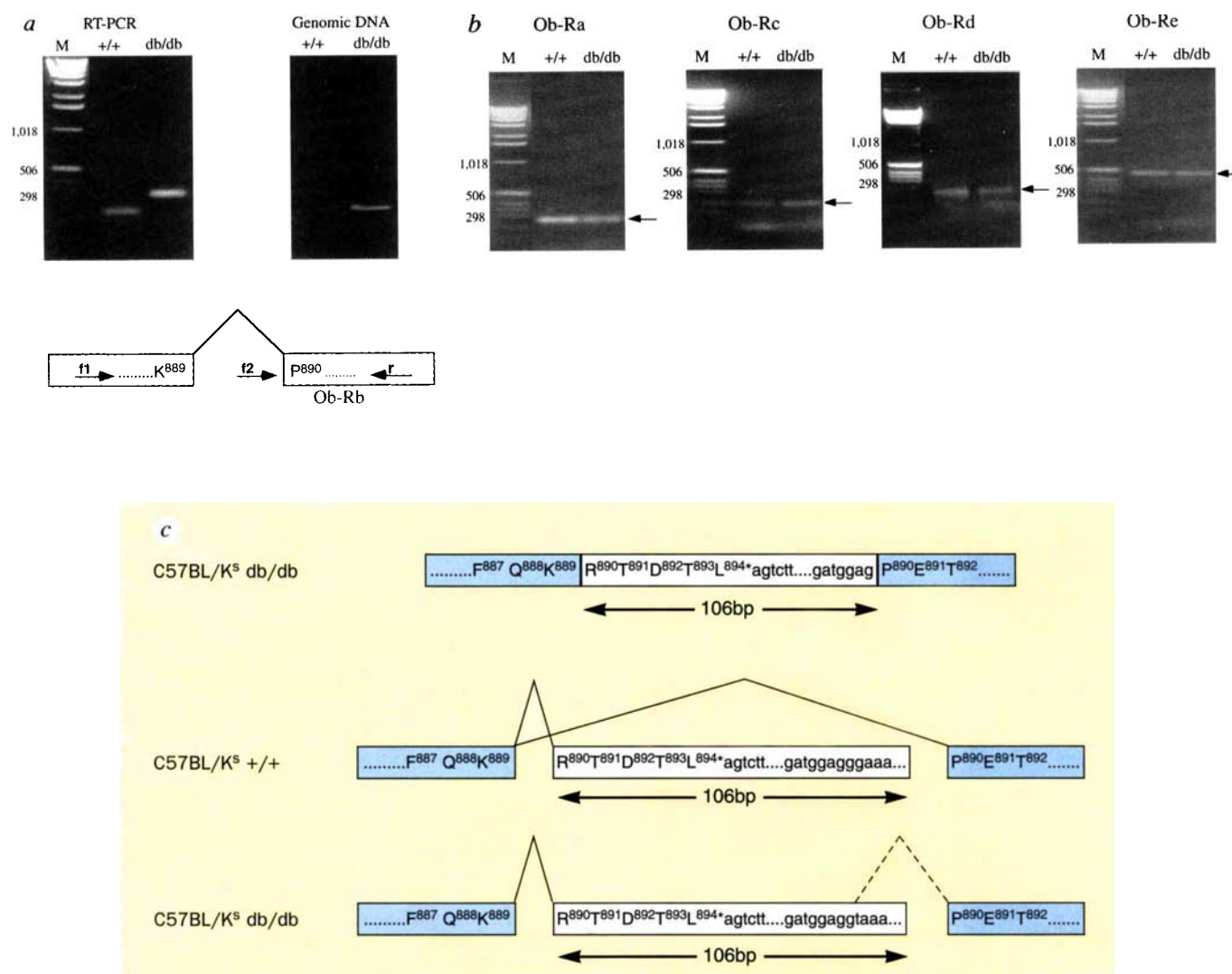


FIG. 3 Abnormal splicing of the leptin receptor in *db* mice. **a**, RT-PCR products from C57BL/K^s *db/db* and wild-type mice were amplified from hypothalamic RNA using a primer pair specific for the Ob-Rb C terminus (see f1 and r; bottom). PCR products of genomic DNA spanning the Ob-Rb splice acceptor in C57BL/K^s *db/db* mice and littermate controls were amplified using primers f2 and r (bottom). **b**, RT-PCR reactions specific for the C termini of Ob-Ra, Ob-Rc, Ob-Rd and Ob-Re were performed using hypothalamic cDNA from *db* and wild-type mice. **c**, DNA sequencing identified a 106-bp insertion in the mutant Ob-Rb RNA at the splice junction between Lys 889 and Pro 890. The sequence of the insertion was identical to the first 106-bp of the C-terminal exon of Ob-Ra (top). The insertion predicts a premature stop codon and changes the amino-acid sequence of Ob-Rb to Ob-Ra. DNA sequencing of the Ob-Ra exon from C57BL/K^s *db/db* and littermate controls revealed a G → T mutation 108 bp after its splice acceptor at Arg 890. This mutation results in the appearance of a consensus splice donor site, AGGTAAA, resulting in the insertion of 106 bp of the C-terminal exon of Ob-Ra into that of Ob-Rb (bottom). **d**, DNA sequencing of the PCR-amplified Ob-Ra C-terminal exon from C57BL/K^s *db/db* mice identified a G → T mutation 108 bp after the splice acceptor at Arg 890.

METHODS. RT-PCR and sequencing were performed as described³. Genomic sequences at the splice acceptor of Ob-Rb were obtained by vectorette PCR of BAC 242 with the Ob-Rb reverse primer. For RT-PCR of Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rd the forward primer was the same 5'-ACACTGTAAATTCACACCAGAG-3' (also labelled f1 in **a**). The reverse primers were: Ob-Ra 5'-AGTCATTCAAACCATTTAGTTAGG-3', Ob-Rb 5'-TGGA-TAAACCCCTTGCTCTCA-3', Ob-Rc 5'-TGAACACAACAATAAAGCCC-3' and Ob-Rd 5'-AGGCTCCCTCAGGCCAC-3'. The intron primer for Ob-Rb (labelled f2 in **a**) was 5'-GTGACTGAATGAAGATGTAATATAC-3'.

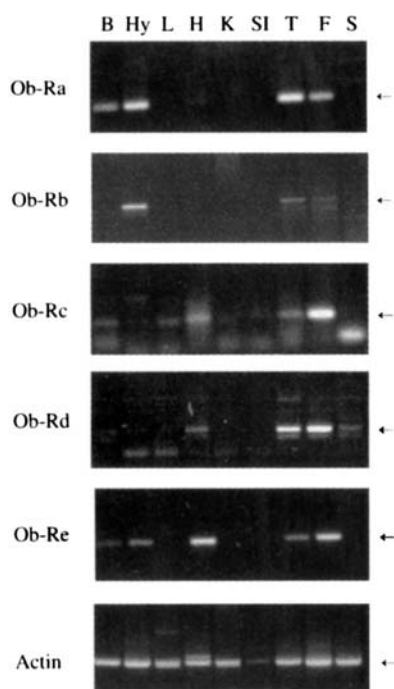


FIG. 4 Tissue distribution of alternatively spliced leptin receptors. RT-PCR for each C-terminal exon was performed from the tissue sources indicated. In each case, one primer from a region of shared nucleotide sequence was used in combination with a primer specific for the alternatively spliced exon. B, brain; Hy, hypothalamus; L, liver; H, heart; K, kidney; SI, small intestine; T, testis; F, adipose tissue; S, spleen. Actin was used as a control. METHODS. RT-PCR was performed as described previously³. The primer sequences for Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rd are shown in the legend to Fig. 3b. The primers for Ob-Re were: 5'-TGTTATATCTGGTTATTGAATGG-3', 5'-CATTAAATGATTATTATCAGAATTGC-3'. The actin primers were used as described previously³.

That the C57Bl/Ks *db/db* mutation is found in the unique C terminus of Ob-Rb explains why the binding of leptin to the choroid plexus was normal in these animals⁸. Taken together, these results suggest that the weight-reducing effects of leptin are at least partly mediated by interactions with an Ob-Rb receptor in the hypothalamus, a brain region known to be important in the regulation of body weight^{4-7,18}. This is supported by the increased potency of leptin when administered directly into the cerebrospinal fluid^{6,7}. Leptin may modulate the activity of NPY, GLP-1 and other peptides that are known to affect feeding behaviour in the hypothalamus and brain^{7,19}. It may also have effects on other tissues expressing the leptin receptor, including adipose tissue. The receptor expressed in choroid plexus, Ob-Ra, may act to transport the protein to the cerebrospinal fluid. Ob-Re, the putative soluble receptor, could function as a transport protein. The functions of the other alternatively processed messages remain to be determined. □

Received 10 January; accepted 24 January 1996.

- Hummel, K. P., Dickie, M. M. & Coleman, D. L. *Science* **153**, 1127–1128 (1966).
- Coleman, D. L. *Diabetologia* **14**, 141–148 (1978).
- Zhang, Y. et al. *Nature* **372**, 425–432 (1994).
- Halaas, J. L. et al. *Science* **269**, 543–546 (1995).
- Pelleymounter, M. A. et al. *Science* **269**, 540–543 (1995).
- Campfield, L. A., Smith, F. J., Guisez, Y., Devos, R. & Burn, P. *Science* **269**, 546–549 (1995).
- Stephens, T. W. et al. *Nature* **377**, 530–534 (1995).
- Tartaglia, L. A. et al. *Cell* **83**, 1263–1271 (1995).
- Bahary, N. et al. *Mamm. Genome* **4**, 511–515 (1993).
- Modi, W. S., Farrar, W. L. & Om, E. *Cytogenet. Cell Genet.* **69**, 232–234 (1995).
- Milatovich, A., Bolger, G., Michaeli, T. & Francke, U. *Somat. Cell molec. Genet.* **20**, 75–86 (1994).
- Church, D. M. et al. *Nature Genet.* **6**, 98–105 (1994).
- Morgan, J. G., Dolganov, G. M., Robbins, S. E., Hinton, L. M. & Lovett, M. *Nucleic Acids Res.* **20**, 5173–5179 (1992).

- Kishimoto, T., Taga, T. & Akira, S. *Cell* **76**, 253–262 (1994).
- Murakami, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11349–11353 (1991).
- Fukunaga, R., Ishizaka-Ikeda, E., Pan, C.-X., Seto, Y. & Nagata, S. *EMBO J.* **10**, 2855–2865 (1991).
- Lodish, H. et al. *Molecular Cell Biology* (Scientific American, New York, 1986).
- Maffei, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6957–6960 (1995).
- Tarton, M. D. et al. *Nature* **379**, 69–72 (1996).
- Bahary, N., Leibel, R. L., Joseph, L. & Friedman, J. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8642–8646 (1990).
- Green, E. D. & Olson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1213–1217 (1990).
- Kusumi, K., Smith, J. S., Segre, J. A., Koos, D. S. & Lander, E. S. *Mamm. Genome* **4**, 391–392 (1993).
- Riley, J. et al. *Nucleic Acids Res.* **18**, 2887–2890 (1990).
- Hermanson, G. G., Hoekstra, M. F., McElligott, D. L. & Evans, G. A. *Nucleic Acids Res.* **19**, 4943–4948 (1991).
- Sternberg, N. *Trends Genet.* **8**, 11–16 (1992).
- Hartl, D. L. & Nurninsky, D. I. *Bio Techniques* **15**, 201–208 (1993).
- Shizuya, H. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8794–8797 (1992).

ACKNOWLEDGEMENTS. We thank S. Burley and J. Froude for comments; S. Korres for assistance in preparing the manuscript; and R. L. Leibel, N. Bahary, R. M. Ramlall-Weiss, R. Schilling and D. McGraw for assistance in characterizing the progeny of the *db × Mus castaneus* cross. This work was supported by a grant from NIH/NIDDK.

Free histidine as a metal chelator in plants that accumulate nickel

Ute Krämer*, Janet D. Cotter-Howells†‡, John M. Charnock†, Alan J. M. Baker§ & J. Andrew C. Smith**||

* Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

† Department of Earth Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

§ Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

A NUMBER of terrestrial plants accumulate large quantities of metals such as zinc, manganese, nickel, cobalt and copper in their shoots¹. The largest group of these so-called 'metal hyperaccumulators' is found in the genus *Alyssum*, in which nickel concentrations can reach 3% of leaf dry biomass^{2,3}. Apart from their intrinsic interest, plants exhibiting this trait could be of value in the decontamination of metal-polluted soils^{4,6}. However, the biochemical basis of the capacity for metal accumulation has not been elucidated. Here we report that exposing hyperaccumulator species of *Alyssum* to nickel elicits a large and proportional increase in the levels of free histidine, which is shown to be coordinated with nickel *in vivo*. Moreover, supplying histidine to a non-accumulating species greatly increases both its nickel tolerance and capacity for nickel transport to the shoot, indicating that enhanced production of histidine is responsible for the nickel hyperaccumulation phenotype in *Alyssum*.

Approximately 400 species of metal-hyperaccumulating plants are currently known, of which about 75% are endemic to ultramafic soils and hyperaccumulate nickel (defined as nickel concentrations exceeding 0.1% of above-ground dry biomass in field samples)¹⁻³. To compare closely related hyperaccumulators and non-accumulators, two species differing greatly in their nickel contents were selected from the genus *Alyssum* (Brassicaceae), which contains 48 hyperaccumulators out of a total of 170 taxa. The hyperaccumulator *A. lesbiacum*^{1,2} was highly tolerant of nickel, showing a 50% growth reduction at about 1.0 mM nickel in the hydroponic solution, more than two orders of magnitude higher than for the non-hyperaccumulator *A. montanum*^{2,7}.

‡ Present address: Department of Plant & Soil Science, University of Aberdeen, Cruickshank Building, St Machar Drive, Aberdeen AB9 2UD, UK.

|| To whom correspondence should be addressed.

(Fig. 1a). This confirms the association between nickel accumulation and nickel tolerance in the genus⁷⁻⁹. The two species differed markedly in their shoot nickel concentrations but not in their root concentrations (Fig. 1b), suggesting that root-to-shoot transport is important in nickel tolerance and hyperaccumulation.

To determine whether selective metal chelation is involved in metal translocation from root to shoot, xylem sap was sampled as root-pressure exudate from the cut surface of de-topped root systems of plants in hydroponic culture. In the hyperaccumulator *A. lesbiacum*, exposure of plants to 0.3 mM nickel caused an increase in the total amino-acid content of the xylem sap. This was almost entirely attributable to a dramatic increase in the concentration of L-histidine, which increased 36-fold from 0.02 to 0.68 mM (Fig. 2a). Exposure of the non-tolerant *A. montanum* to 0.003 mM nickel did not cause any significant change in xylem amino acids (Fig. 2a). Moreover, these nickel treatments did not elicit changes in the principal xylem organic-acid anions in either species (see legend to Fig. 2a).

Over a wide range of nickel treatments, there was a linear relationship between xylem nickel and histidine concentrations in *A. lesbiacum*. Two other hyperaccumulator species, *A. murale* and *A. bertolonii*, also followed this general relationship (Fig. 2b). Of all the organic acids and proteinaceous amino acids, histidine has the highest association constant¹⁰ for complex formation with nickel, both at the observed xylem-sap pH of 5.72 ± 0.08 (mean $\pm$ s.e.m., $n = 18$) and at typical cytosolic pH values around 7.5 (ref. 11). Although organic acids such as citrate, malate and malonate participate in nickel chelation in the shoots of hyperaccumulators^{2,12,13}, they are present constitutively at high concentrations in most terrestrial plants¹⁴ and account for neither the metal specificity nor the species specificity of hyperaccumulation^{3,15-17}. Nickel-hyperaccumulator species of *Alyssum*

also have a limited capacity to hyperaccumulate cobalt in their shoots, although this metal is considerably more toxic than nickel^{9,18}. Exposure of *A. lesbiacum* to low concentrations of cobalt resulted in a linear increase in xylem histidine with increasing cobalt (inset to Fig. 2b), indicating that the histidine response can be elicited by related metal ions.

The apparent importance of histidine in metal hyperaccumulation in *Alyssum* (Fig. 2) is in striking agreement with a prediction¹⁷ that nickel uptake by such plants should involve a nitrogen-containing ligand. Indeed, this principle is exploited in metal-affinity chromatography for purification of recombinant proteins expressing a six-histidine tag on a matrix of immobilized nickel¹⁹. The quantitative difference between the nickel-histidine and cobalt-histidine relationships (Fig. 2b) may be a consequence of the considerably greater stability of the nickel complex¹⁰. Modelling of the metal-ligand complexes present in *A. lesbiacum* xylem exudates using GEOCHEM-PC (ref. 20) indicated that, in plants grown in 0.3 mM nickel, $97.7 \pm 0.8\%$ (mean $\pm$ s.e.m., $n = 4$) of the histidine was complexed with nickel (the remaining 2.3% simply being protonated histidine). Thus, essentially all the histidine produced in these plants is complexed with the metal ion that induced its synthesis.

To obtain direct information on the coordination chemistry of nickel in *A. lesbiacum*, samples were analysed by extended X-ray absorption fine structure (EXAFS) spectroscopy, a technique applicable to biological materials that lack long-range order^{21,22}. EXAFS spectra were collected for xylem sap of *A. lesbiacum* (Fig. 3a). The radial distribution function around the nickel atom (Fig. 3b) indicated a shell of backscattering atoms at roughly 2 Å, with further shells at roughly 3 Å and 4.2 Å. A best fit to these data was given by octahedral coordination of nickel with an imidazole nitrogen of histidine at 1.98 Å and five other light atoms at 2.05 Å

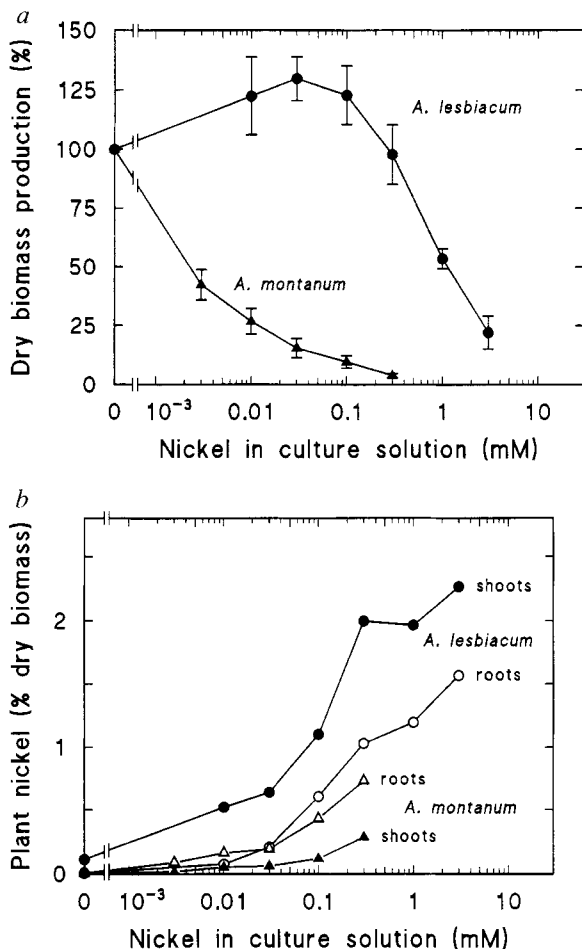


FIG. 1 Nickel tolerance and nickel uptake in the hyperaccumulator *Alyssum lesbiacum* (Candargy) Rech. f. and the non-hyperaccumulator *A. montanum* L. a, Whole-plant biomass production expressed as percentage of controls grown in the absence of nickel (dry biomass per plant for the controls averaged 27.4 ± 3.1 mg for *A. lesbiacum* and 29.0 ± 4.0 mg for *A. montanum*). Values are means $\pm$ s.e.m. of four experiments for *A. lesbiacum* and of two experiments for *A. montanum* ($n = 10$ plants per treatment). b, Nickel concentrations in dry biomass of roots and shoots. Values are means of two independent experiments, in each of which material was pooled from 10 plants.

METHODS. Plants were germinated for 7 d on floating nets in 1 mM $\text{Ca}(\text{NO}_3)_2$ solution, subsequently transferred to 1.2-l culture vessels (10 plants per vessel), and supplied with modified 0.1-strength Hoagland solution containing 0.1 mM KH_2PO_4 , 0.5 mM KNO_3 , 0.4 mM $\text{Ca}(\text{NO}_3)_2$, 0.2 mM MgSO_4 , 0.01 mM FeNaEDTA , 0.01 mM H_3BO_3 , 2 μM MnSO_4 , 0.2 μM ZnSO_4 , 0.2 μM CuSO_4 , 0.1 μM Na_2MoO_4 and 0.02 mM NaCl , supplemented with 0, 0.003, 0.01, 0.03, 0.1, 0.3, 1, or 3 mM NiSO_4 . Plants were grown in a glasshouse at night temperatures between 14 °C and 18 °C and day temperatures between 22 °C and 34 °C with supplementary lighting provided by sodium-vapour lamps to give a 16-h photoperiod. Culture solutions were continuously aerated and exchanged every 7 d; solution pH remained in the range 5.6–6.0. Plants were harvested after 18 d, rinsed with distilled water, dried at 80 °C for 48 h and weighed. For analysis, duplicate samples of dried plant material were ashed in a muffle furnace (16 h at 400 °C, followed by 24 h at 500 °C), dissolved in 20% (v/v) HCl , and nickel quantified in diluted solutions containing 1% (v/v) H_2SO_4 and 0.04% (w/v) LaCl_3 using a Perkin Elmer 3030 atomic absorption spectrophotometer.

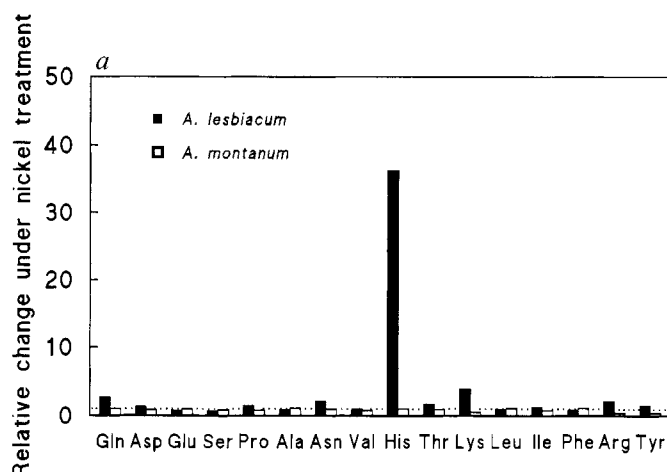
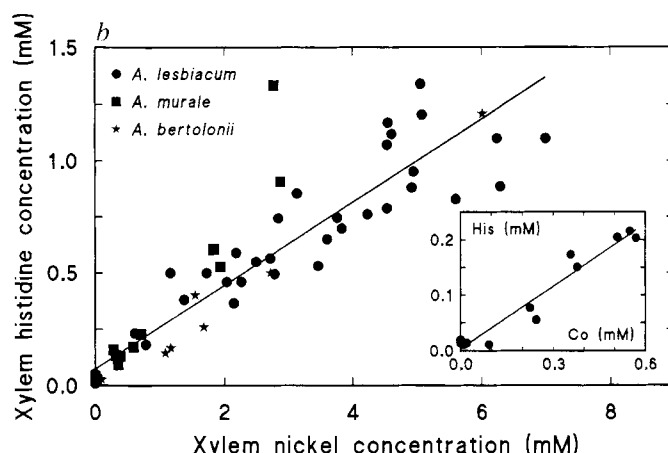


FIG. 2 Changes in amino-acid composition of xylem sap as a response to nickel. *a*, Ratios of xylem amino-acid concentrations in plants exposed to nickel in the root medium (0.3 mM NiSO_4 for *A. lesbiacum*, 0.003 mM for *A. montanum*) relative to amino-acid concentrations in the xylem sap of control plants (0 mM NiSO_4). Ratios were calculated from the means of values from four independent batches of plants, within each of which five plants were pooled for analysis. Amino acids are listed from left to right in order of decreasing abundance in the xylem sap of *A. lesbiacum* controls; only those amino acids present at ≥ 0.002 mM in any sample are shown. The dotted line indicates a ratio of 1. The only significant change observed in xylem amino-acid concentrations for either species was for histidine in *A. lesbiacum* ($P < 0.001$; two-sided *t*-test of primary data for controls versus nickel-treated plants). The principal xylem organic-acid anions determined by high-performance liquid chromatography (HPLC) using a Bio-Rad HPX-87H ion-exclusion column were citrate and malate, but neither changed significantly in either species on exposure to nickel (citrate and malate concentrations, assayed according to ref. 30, averaging 0.30 mM and 0.15 mM, respectively, in *A. lesbiacum*, and 0.20 mM and 0.05 mM, respectively, in *A. montanum*). Modelling of nickel speciation in the xylem sap of *A. lesbiacum* using GEOCHEM-PC version 2.0 (ref. 20) indicated that nickel was chelated mainly by histidine (19%), glutamine (15%), citrate (9%) and malate (3%), with up to 48% predicted to be the free (hydrated)



Ni^{2+} ion. *b*, Relationship between nickel and histidine concentrations in xylem sap of *A. lesbiacum* and two other hyperaccumulator species (*A. murale* Waldst. & Kit. and *A. bertolonii* Desv.). Regression analysis yielded the equation $y = 0.186x + 0.072$ ($r^2 = 0.851$, d.f. = 63, $P < 0.001$). The inset shows the equivalent relationship between xylem cobalt and histidine concentrations for *A. lesbiacum*, fitted by the equation $y = 0.376x + 0.003$ ($r^2 = 0.945$, d.f. = 10, $P < 0.001$). Data were obtained through analysis of root-pressure exudates obtained as xylem sap from de-topped root systems of plants grown at a range of nickel or cobalt concentrations (0–3 mM NiSO_4 , or 0–0.03 mM CoSO_4 , added to the nutrient solution described in Fig. 1).

METHODS. Plants were cultivated after germination (7 d) for 4 weeks as in Fig. 1 and were then transferred to nickel-containing solutions for 8 d before sampling. To obtain xylem sap, shoots were excised 0.5 h after the onset of the dark period, the cut surfaces of the hypocotyls blotted with absorbent tissue, and root-pressure exudate collected over 8 h. Samples were analysed for inorganic cations by atomic absorption spectrophotometry. Amino acids were analysed using an Applied Biosystems PTC C-18 reverse-phase column (220 $\times$ 2.1 mm) in an ABI 420A derivatizer/analyser after pre-column derivatization with phenylisothiocyanate; amino acids were separated at 34 $^\circ\text{C}$ by applying a sodium acetate/acetonitrile gradient in water.

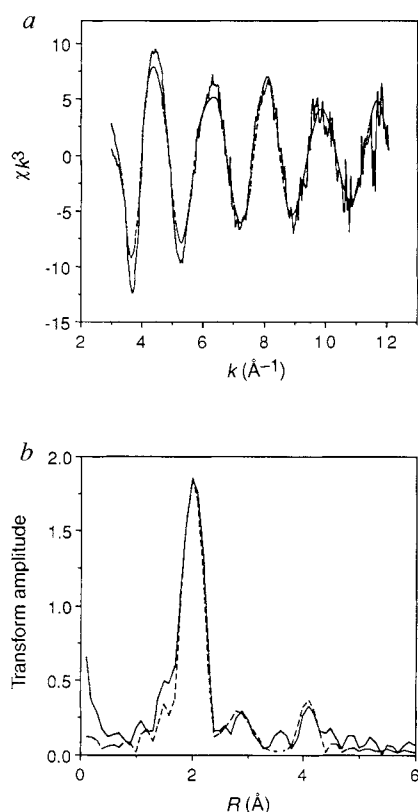
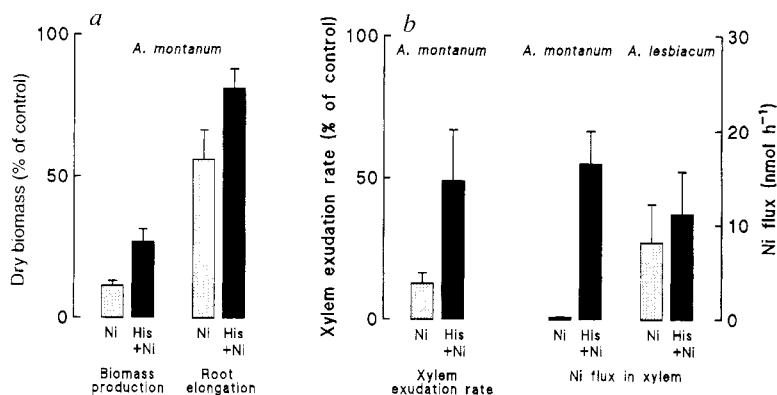


FIG. 3 Extended X-ray absorption fine structure (EXAFS) analysis of nickel in xylem sap from *A. lesbiacum*. *a*, Normalized oscillatory EXAFS amplitude (χ), weighted by k^3 , plotted against the photoelectron wave vector (k). *b*, Associated Fourier transform (R is the distance of scattering atoms from the primary absorber). Continuous lines are experimental data and broken lines the best theoretical fit.

METHODS. Xylem sap, collected after exposing plants to 0.3 mM Ni in the root medium for 8 d as described in Fig. 2, was injected into a perspex sample cell with Mylar windows and frozen in liquid nitrogen. EXAFS data were collected at 77 K in fluorescence mode at the nickel *K*-edge using a Canberra 13-element solid-state detector on station 8.1 at the Daresbury Synchrotron Radiation Source, operating at 2 GeV with an average current of 150 mA. Eight scans were averaged and the spectra analysed, including multiple scattering, using the Daresbury program EXCURV92; theoretical fits were generated by adding shells of scatterers around the central nickel atom and iterating bond lengths and Debye–Waller type factors to obtain the best fit to the experimental data²².

FIG. 4 Effect of exogenously applied histidine on nickel tolerance and xylem nickel translocation. **a**, Effect of foliar spraying with histidine (compared with glutamine) on nickel tolerance of *A. montanum*, assayed as the effect on whole-plant dry biomass production and root elongation. After germination for 7 d, plants were grown in modified Hoagland solution (Fig. 1) containing either 0 or 0.01 mM NiSO_4 for 23 d and were sprayed daily with a solution of either 20 mM glutamine (Ni) or 20 mM histidine (His + Ni), both containing 0.01% (v/v) of the non-ionic detergent Tween 20. Values of biomass production and root elongation are for nickel-treated plants expressed as percentages of the respective controls grown in the absence of nickel. Each bar represents the mean of 9 plants + 1 s.e.m. Biomass production ($P < 0.001$) and root elongation ($P = 0.010$, both one-sided *t*-tests) were significantly greater in nickel-treated plants sprayed with histidine compared with glutamine. **b**, Effect of histidine supplied in the root medium on xylem-sap exudation rate and on nickel flux in *A. montanum* and *A. lesbiacum* exposed to 0.3 mM NiSO_4 . Roots of 4-week-old plants were excised and exposed to a root medium of modified Hoagland solution supplemented with either 0.3 mM MgSO_4 (controls) or 0.3 mM NiSO_4 , with or without 0.3 mM added histidine. Xylem exudate was collected over 12 h. Each bar represents the geometric



mean of 10 samples + 1 s.e.m. For *A. montanum*, both xylem exudation rate and nickel flux were significantly greater at $P < 0.001$ in the presence of histidine (ANOVA); no significant differences were observed in *A. lesbiacum* (exudation rates not shown).

(mean square variation represented by Debye–Waller factors of $< 0.001 \text{ \AA}^2$ and $0.004 \text{ \AA}^2$, respectively). The smaller contributions from outer-shell carbon and nitrogen atoms at 2.93 Å and 4.24 Å (Debye–Waller factors $0.004 \text{ \AA}^2$ and $0.008 \text{ \AA}^2$, respectively) are characteristic of imidazole ring coordination with intensities enhanced by multiple scattering²³. Almost identical spectra were obtained for whole leaves and roots from the same plants (data not shown), in which histidine levels were elevated 20-fold and 3-fold, respectively, by exposure to nickel.

These results demonstrate that nickel is complexed with histidine in tissues of *A. lesbiacum*. There was no indication of nickel coordination by sulphur, for which a Ni–S distance of roughly 2.25 Å would be predicted. This is significant because several current theories of metal tolerance in plants implicate metallothioneins, which include cysteine-rich proteins and low-molecular-weight polypeptides (phytochelatins), in the complexing of metal ions^{24–26}. In particular, these ligands can reduce the toxicity of cadmium and copper in plants^{26,27}, but their synthesis often does not correlate with metal tolerance^{28,29}. Production of a free amino acid or organic acid in plants as a specific and proportional response to metal treatment (Fig. 2) has not to our knowledge been previously observed.

To test the hypothesis that histidine production is responsible for nickel tolerance in hyperaccumulators, we supplied the non-tolerant *A. montanum* with histidine as a foliar spray (glutamine

being used for comparison as an amino acid with similar physico-chemical properties). At toxic concentrations of nickel, histidine application more than doubled plant biomass production and approximately halved the inhibitory effect of nickel on root elongation (Fig. 4a). Moreover, supplying histidine in the root medium moderated the adverse effect of nickel on xylem exudation and greatly increased nickel flux through the xylem of *A. montanum* (Fig. 4b), while having no significant effect on the nickel-tolerant *A. lesbiacum*. These results indicate that histidine is involved both in the mechanism of nickel tolerance and in the high rates of nickel transport into the xylem required for hyper-accumulation in the shoot.

Since the discovery of metal hyperaccumulation in plants, it has been assumed that intracellular detoxification of the metal is achieved by production of an appropriate high-affinity ligand. Nickel detoxification is probably of primary importance in the root-cell cytoplasm rather than the xylem, as the latter represents an apoplastic, extracellular phase, but effective translocation of nickel from root to shoot in the xylem is essential for generating the hyperaccumulator phenotype. Our study suggests that the amino acid histidine plays a role in both of these processes in *Alyssum*. An improved understanding of the mechanisms underlying hyperaccumulation of nickel and other metals should open up new possibilities for the application of metal-accumulating plants in the phytoremediation^{4–6} of contaminated soils. □

Received 1 September; accepted 19 December 1995.

- Baker, A. J. M. & Brooks, R. R. *Biorecovery* **1**, 81–126 (1989).
- Brooks, R. R., Morrison, R. S., Reeves, R. D., Dudley, T. R. & Akman, Y. *Proc. R. Soc. Lond. B* **203**, 387–403 (1979).
- Reeves, R. D. in *The Vegetation of Ultramafic (Serpentine) Soils* (eds Baker, A. J. M., Proctor, J. & Reeves, R. D.) 253–277 (Intercept, Andover, 1992).
- Chaney, R. L. in *Land Treatment of Hazardous Wastes* (eds Parr, J. F., Marsh, P. B. & Kila, J. M.) 50–76 (Noyes Data Corp., Park Ridge, NJ, 1983).
- Baker, A. J. M., McGrath, S. P., Sidoli, C. M. D. & Reeves, R. D. *Resources Conserv. Recycling* **11**, 41–49 (1994).
- Salt, D. E. et al. *Biotechnology* **13**, 468–474 (1995).
- Morrison, R. S., Brooks, R. R. & Reeves, R. D. *Pl. Sci. Lett.* **17**, 451–457 (1980).
- Gabbriellini, R., Pandolfini, T., Vergnano, O. & Palandri, M. R. *Pl. Soil* **122**, 271–277 (1990).
- Homer, F. A., Morrison, R. S., Brooks, R. R., Clemens, J. & Reeves, R. D. *Pl. Soil* **138**, 195–205 (1991).
- Smith, R. M. & Martell, A. E. *Critical Stability Constants* Vol. 6; second suppl. (Plenum, New York, 1989).
- Smith, F. A. & Raven, J. A. *Rev. Pl. Physiol.* **30**, 289–311 (1979).
- Pancaro, L., Pelosi, P., Vergnano Gambi, O. & Galoppini, C. *Giorn. Bot. Ital.* **112**, 141–146 (1978).
- Brooks, R. R., Shaw, S. & Asensi Marfil, A. *Physiologia Pl.* **51**, 167–170 (1981).
- Osmond, C. B. in *Encyclopedia of Plant Physiology (New Series)* Vol. 2B (eds Lüttge, U. & Pitman, M. G.) 346–372 (Springer, Berlin, 1976).
- Woolhouse, H. W. in *Encyclopedia of Plant Physiology (New Series)* Vol. 12C (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Ziegler, H.) 245–300 (Springer, Berlin, 1983).
- Ernst, W. H. O., Verkleij, J. A. C. & Schat, H. *Acta Bot. Neerl.* **41**, 229–248 (1992).
- Still, E. R. & Williams, R. J. P. *J. inorg. Biochem.* **13**, 35–40 (1980).

- Gabbriellini, R., Mattioni, C. & Vergnano, O. J. *Pl. Nutr.* **14**, 1067–1080 (1991).
- Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. & Stuber, D. *Biotechnology* **6**, 1321–1325 (1988).
- Parker, D. R., Norvell, W. A. & Chaney, R. L. in *Chemical Equilibrium and Reaction Models* (eds Loeppert, R. H., Schwab, A. P. & Goldberg, S.) 253–269 (S.S.S.A. Special Publication Number 42, Soil Science Society of America/American Society of Agronomy, Madison, Wisconsin, 1995).
- Königsberger, D. & Prins, R. (eds) *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES* (Wiley, New York, 1988).
- Chamock, J. M. *Radiat. Phys. Chem.* **45**, 385–391 (1995).
- Blackburn, N. J. et al. *Biochem. J.* **213**, 765–768 (1983).
- Grill, E., Winnacker, E.-L. & Zenk, M. H. *Science* **230**, 674–676 (1985).
- Rauscher, W. E. A. *Rev. Biochem.* **59**, 61–86 (1990).
- Robinson, N. J., Tommey, A. M., Kuske, C. & Jackson, P. J. *Biochem. J.* **295**, 1–10 (1993).
- Howden, R., Goldsbrough, P. B., Andersen, C. R. & Cobbett, C. S. *Pl. Physiol.* **107**, 1059–1066 (1995).
- Delhaize, E., Jackson, P. J., Lujan, L. D. & Robinson, N. J. *Pl. Physiol.* **89**, 700–706 (1989).
- de Knecht, J. A. et al. *Pl. Physiol.* **104**, 255–261 (1994).
- Bergmeyer, H. U. (ed.) *Methods of Enzymatic Analysis* 3rd edn Vol. VII (VCH, Weinheim, 1985).

ACKNOWLEDGEMENTS. We thank T. Pandolfini and R. Gabbriellini for seed of *Alyssum bertolonii*, A. C. Willis and K. B. M. Reid for amino-acid analysis, J. F. W. Mosselmans for technical support at the Daresbury Laboratory, and P. R. Darrah, C. J. Leaver and C. K. Prout for discussions. This work was supported by the Rhodes Trust, a NERC postdoctoral fellowship to J.D.C.-H., and the Council for the Central Laboratory of the Research Councils.

Amplification of high-frequency synaptic inputs by active dendritic membrane processes

Juergen Haag & Alexander Borst*

Friedrich-Miescher-Laboratorium der Max-Planck-Gesellschaft,
Spemannstrasse 37-39, D-72076 Tuebingen, Germany

*To whom correspondence should be addressed.

ACTIVE membrane mechanisms have been found in the dendrites of many nerve cells¹⁻⁷. Their contribution to dendritic signal processing, however, remains unclear⁸ because few experimental preparations allow for detailed characterization of dendritic physiology under normal conditions. To investigate the functional implications of active dendritic processes in an *in vivo* preparation, we compared the response properties of different types of non-spiking, motion-sensitive interneurons of the fly visual system, only one of which is equipped with a fast sodium inward current in its axon as well as in its dendrite. We report here that cells with fast activating sodium currents can respond to temporal changes in their synaptic input signals up to much higher frequencies than can those which lack such currents. Thus fast sodium currents lead to a frequency-dependent amplification of synaptic signals, enhancing cellular responses specifically to transient inputs which otherwise would be attenuated because of the passive properties of the dendritic tree.

The large, motion-sensitive visual interneurons ('tangential cells') in the lobula plate of the blowfly *Calliphora erythrocephala*^{9,10} represent a system in which the function of voltage-gated channels in graded signal transmission can be addressed experimentally *in vivo*. Each of the roughly 60 anatomically distinct and individually identifiable tangential cells receives input from several hundred columnar, retinotopically arranged elements¹¹⁻¹³ on an extended dendritic tree. Although this set of inputs is common to all tangential neurons, their output signals reveal three physiologically different subsets of neurons. When stimulated by visual motion, some of the tangential cells (such as the H1 cell) respond with an increased firing frequency, some

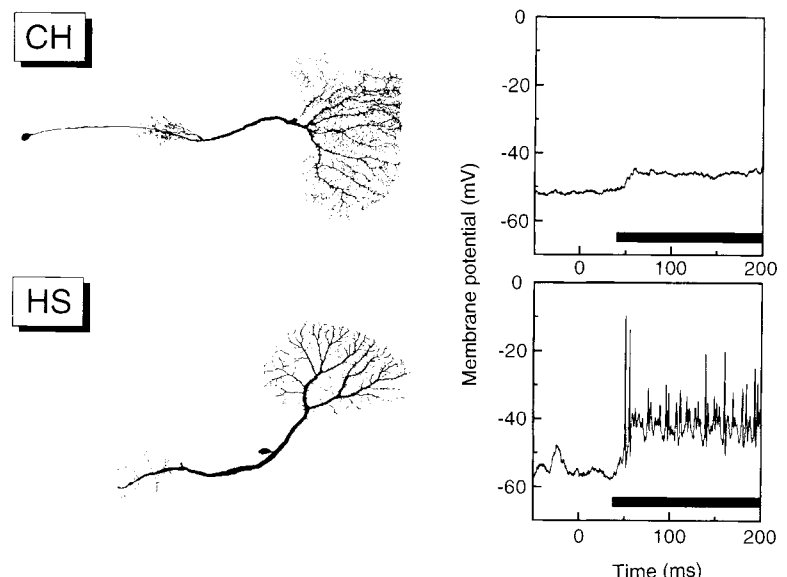
respond only with a shift of their steady-state membrane potential (such as the CH cells), and still others respond with irregular spike-like events superimposed on a graded membrane potential shift (such as HS and VS cells)¹⁴. These properties suggest large differences with respect to the active membrane properties of the different groups of neurons, thereby providing a unique opportunity to analyse the functional implications of voltage-gated ion channels.

The experiments presented here were done on two types of tangential cells, the CH cells¹⁵⁻¹⁷ and the HS cells^{18,19} (Fig. 1). Switched electrode voltage-clamp experiments reveal the ionic basis for their different visual response properties²⁰: while CH cells possess only a slow calcium inward current and a delayed rectifying potassium outward current, HS cells exhibit, in addition, a fast activating and rapidly inactivating sodium inward current. As is demonstrated by dual electrode recordings from HS cells, the fast sodium current is present not only in the axon of these cells but in their major dendritic branches as well (Fig. 2). Model simulations show that the experimentally determined characteristics of the sodium and potassium activation and inactivation curve (along with the additional assumption of a small maximum conductance value) give rise to the irregular spike-like events observed in real HS-cells (Fig. 1) when depolarized by either synaptic stimulation or current injection²⁰.

A significant physiological manifestation of fast inward currents is seen in the cells' membrane responses to current injections of various frequencies (Fig. 3). When a sinusoidal current is superimposed on a negative d.c. signal of -4 nA, the voltage responses of both HS and CH cells behave as typical low-pass filters in their amplitude spectrum: HS cells have a cutoff frequency of about 80 Hz (Fig. 3b, filled circles), compared to about 50 Hz in CH cells (Fig. 3a, filled circles). Because the negative d.c. injection shifts the membrane potential to a range in which no voltage-gated currents are activated, these spectra reflect the passive membrane properties of both cell types. In contrast, when the sinusoidal current modulation is done with frequencies above 10 Hz without applying a negative d.c. current, an amplification of the HS cell response is observed (Fig. 3b, open circles). In this frequency range, spikes are elicited regularly in HS cells at the positive peaks of the sinusoidal current: although the inward currents are normally inactive at the resting potential, they become activated when inactivation is removed by the short hyperpolarization

FIG. 1 Intracellularly recorded responses to motion stimuli (indicated by the black bar) as recorded under current-clamp conditions from the axons of two different classes of lobula plate tangential cells, CH and HS cells. Both CH and HS cells are excited by horizontal motion from the front to the back within the visual field of the ipsilateral eye (with respect to their lobula plate arborization) and are inhibited by motion in the opposite direction. A small number of cells of each type is found per lobula plate differing mainly with respect to the vertical location of their receptive field: two CH cells (the dorsal 'dCH' cell and the ventral 'vCH' cell) and three HS cells (the northern 'HSN', the equatorial 'HSE' and the southern 'HSS' cell). Each cell's receptive field location closely coincides with the extent of its dendritic tree within the lobula plate. While CH cells respond to visual patterns moving in the preferred direction with a purely graded shift of their membrane potential, HS cells have additional spike-like events superimposed onto the graded response. To the left, the anatomy of each cell type is shown as reconstructed from cobalt-filled material.

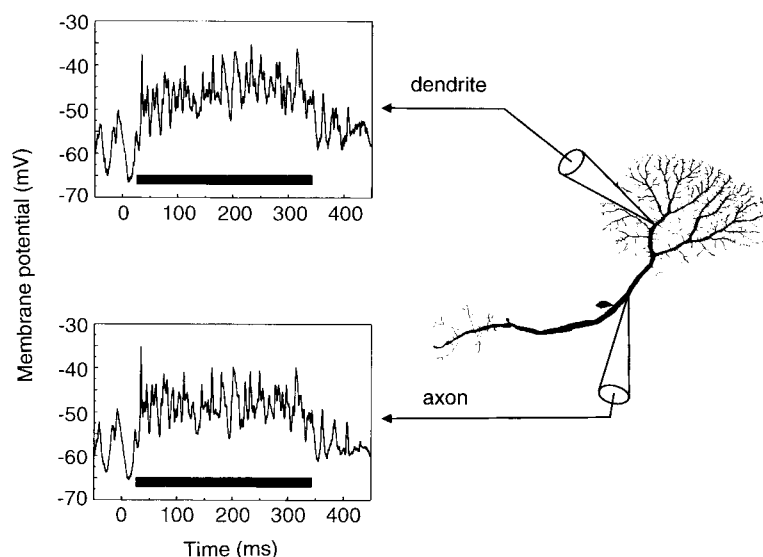
METHODS. For details of the dissection procedure, see ref. 13. Stimuli were generated on Tektronix 608 monitors by an image synthesizer (Picasso, Innisfree) and consisted of a one-dimensional square-wave grating of 14° spatial wavelength and 87% contrast displayed at a frame rate of 200 Hz. The angular width of the stimulus fields was 40° in the horizontal and 28° in the vertical direction as seen by the fly. When activated, the grating was moving at 28° s⁻¹. Recording electrodes were made of glass capillaries (Clark Electromedical, GC100F-10) pulled on a P80 Brown-Flaming puller;



when filled with 1 M KCl they had a resistance of about 20 MΩ. Signals were amplified (SEC-10L, npi-electronics) and fed to a PC (IBM PS/2) through an A/D converter (Metra Byte μCDAS-16G, Keithley Instruments) at 10 kHz.

FIG. 2 Simultaneous dual-electrode recording of the motion response from the axon and the dendrite of an HS cell under current-clamp conditions. The motion stimulus is indicated by the black bar underneath each response trace. Note that the fast and irregular spike-like events are found with identical amplitude in the dendrite as well as in the axon. Model simulations show that the dendritic signals are not compatible with an electrically passive dendritic membrane.

METHODS. For dual-electrode recordings the fly was mounted horizontally underneath a microscope (AxioTech Vario, Zeiss) looking down onto the stimulus monitor. With the first electrode, filled with a fluorescent dye (Calcium Green, Molecular Probes), the cell was penetrated in the axon and stained by injection of hyperpolarizing current pulses (50% duty cycle, 2 s cycle duration, 4 nA amplitude) for several minutes. The second electrode was then placed into the dendrite under visual control. All other experimental parameters are as specified in the legend to Fig. 1.



phases ('rebound spikes'). As can be seen in the amplitude spectra of HS cells, these inward currents serve to amplify the neuron's response to inputs varying sinusoidally within a certain frequency range. Under the same conditions, the CH cell response remains unaltered (Fig. 3a, open circles).

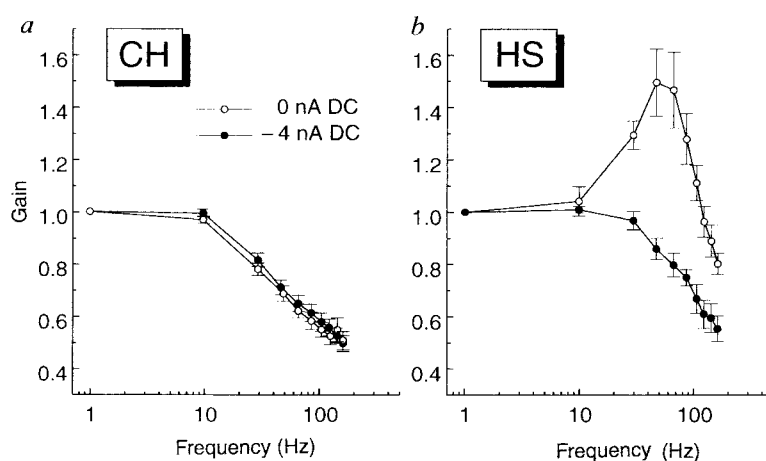
Does this property of HS cells play any role when the responses to their natural synaptic input signals are considered, that is, when displaying motion stimuli in front of the fly's eye instead of injecting current? To address this question we again recorded intracellularly from these neurons and presented motion stimuli, the velocity of which were modulated sinusoidally within the same frequency range as was done before with the current injections. HS cells (Fig. 4b) indeed respond to high-velocity modulations much better than do CH cells (Fig. 4a). However, there are several response features that require further explanation: (1) the responses of both cell types differ significantly from a sinusoid; (2) the responses are damped at much lower frequencies than is the case when current is injected into the axon; and (3) the high-frequency responses exhibit a transient behaviour. Many of these points can be explained by the different site of signal generation (dendrite versus axon) and by the fact that synaptic input signals lead to conductance changes which, in contrast to current injection, are related to the membrane potential in a nonlinear way²¹.

Furthermore, the properties of the input elements have to be taken into account. These are local motion-sensitive units with a nonlinear dependence on image velocity which respond to changes in image velocity of increasing frequency with a decreased amplitude²²⁻²⁴. The responses obtained from the tangential cells under motion stimulation thus reflect a superposition of both their intrinsic membrane characteristics and the dynamic response properties of the incoming fibres.

Because the responses of HS cells to high-frequency stimuli are not centred around resting potential but consist mainly of depolarizing events, it is unlikely that different passive membrane properties of HS and CH cells are responsible for the differences in their response patterns to visual motion. An even more direct way of demonstrating the contribution of voltage-gated currents to the normal motion response of HS cells is to inject a large depolarizing current into the cells, which inactivates voltage-gated sodium currents, while stimulating them synaptically with visual motion. Figure 4c shows the result of such an experiment in which a d.c. current of +4 nA was injected into an HS cell. At a frequency of 1 Hz, the response is still similar to that found without current injection. However, at stimulus frequencies above 20 Hz, the response amplitudes are much reduced. Thus, with the inward currents inactivated, HS cell responses become more like those of

FIG. 3 Amplitude spectra of CH (a) and HS cells (b) obtained by injection of sinusoidal currents of various frequencies into the axon. The filled circles represent the voltage response when the sinusoidal current modulation is superimposed on a negative d.c. current of -4 nA; the open circles represent the responses when no additional d.c. current is applied and, thus, voltage modulation is around the resting potential of the cells. Data are the mean values $\pm$ s.e.m. obtained from 3 HS cells and 10 CH cells, respectively. In CH cells, spectra with and without additional d.c. current are similar. In contrast, in HS cells there is a considerable amplification of the response amplitude in the frequency range above 10 Hz when modulation occurs around the resting potential of the cell, in comparison with the spectrum when the cell is hyperpolarized. Similar spectra can be obtained from computer simulations taking into account the detailed anatomy of the cells and the properties of their voltage-gated ion channels.

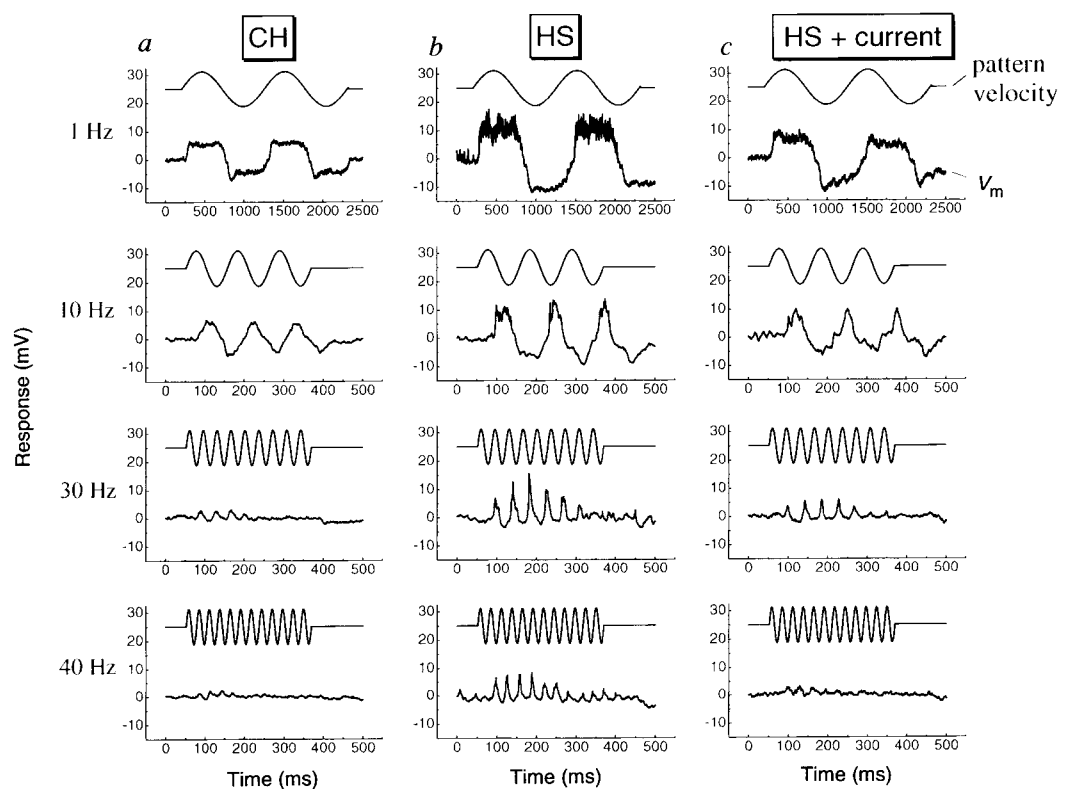
METHODS. Using the current-clamp mode of the amplifier (SEC-10L, npi-electronics), sinusoidal current of 1.2 nA amplitude was injected in sweeps of 0.5 s (10 Hz and higher) or 2.5 s length (1 Hz) going through frequencies of 1, 10, 30 ... 200 Hz into the cell. Each injection was repeated 10 times per frequency with an intermitting time span of 1 s. Sweeps were averaged for each frequency off-line. From these data, Fourier spectra were calculated using the FFT algorithm for both the



injected current and the voltage response. For each record the resulting voltage spectrum was divided by that of the injected current. The resulting input resistance at the stimulus frequency was normalized with respect to the value obtained for 1 Hz before averaging.

FIG. 4 Responses of CH (a) and HS cells (b, c) when stimulated by a visual pattern displayed within the receptive field of the cells. The pattern was moving with a sinusoidally modulated velocity at frequencies of 1 to 40 Hz (from top to bottom). The temporal velocity profile is displayed within each panel on top of the voltage response V_m of the cells. In c, HS cells were additionally manipulated by injection of a constant depolarizing current of +4 nA to inactivate voltage-gated sodium currents. Note that, without additional manipulation, HS cells follow pattern motion to much higher frequencies than CH cells (compare a and b). With the depolarizing current injected (c), HS cells' responses at high stimulus frequencies are much reduced and, again, become more like those of CH cells. The membrane potential is not given in absolute values but relative to the potential before stimulus onset.

METHODS. Sinusoidal velocity modulation of the grating was displayed with frequencies ranging from 1 to 40 Hz. Maximum pattern displacement was reduced with increasing stimulus frequency to keep maximum velocity constant at 120°s^{-1} , for all frequencies. Each stimulus condition was repeated 10 times for each cell with an intermitting



span of 1 s during which time the pattern was at rest. All other parameters of the visual stimulus were the same as specified in the legend to Fig. 1. Similar data were obtained in 3 other CH cells and 5 other HS cells.

CH cells, substantiating the causal relationship between the voltage-gated sodium current and the high-frequency responses of the HS cells. One obvious function of the fast inward current observed in HS but not in CH cells, therefore, is to allow HS cells to transmit information about fast changes in image motion from

the postsynaptic sites in the dendrite to the axonal terminals located in the protocerebrum. Such an amplification of high-frequency signals may not be necessary in CH cells because their output synapses are found within the dendrite in close vicinity to their synaptic input sites²⁵. □

Received 30 August; accepted 3 November 1995.

- Spencer, W. A. & Kandel, E. R. *J. Neurophysiol.* **24**, 272–285 (1961).
- Stuart, G. J. & Sakmann, B. *Nature* **367**, 69–72 (1994).
- Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. *Science* **268**, 297–300 (1995).
- Magee, J. C. & Johnston, D. *Science* **268**, 301–304 (1995).
- Llinas, R. *Science* **242**, 1654–1664 (1988).
- Laurent, G. *J. Physiol., Lond.* **470**, 45–54 (1993).
- Yuste, R., Gutnick, M. J., Saar, D., Delaney, K. R. & Tank, D. W. *Neuron* **13**, 23–43 (1994).
- Bernander, O., Koch, C. & Douglas, R. J. *J. Neurophysiol.* **72**, 2743–2753 (1994).
- Hausen, K. in *Photoreception and Vision in Invertebrates* (ed. Ali, M. A.) 523–559 (Plenum, New York, 1984).
- Egelhaaf, M. & Borst, A. *J. Neurosci.* **13**, 4563–4574 (1993).
- Strausfeld, N. J. in *Photoreception and Vision in Invertebrates* (ed. Ali, M. A.) 483–522 (Plenum, New York, 1984).
- Borst, A. & Egelhaaf, M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4139–4143 (1992).
- Haag, J., Egelhaaf, M. & Borst, A. *Neurosci. Lett.* **140**, 173–176 (1992).
- Hengstenberg, R. *Nature* **270**, 338–340 (1977).
- Hausen, K. *Z. Naturforsch.* **31c**, 629–633 (1976).
- Eckert, H. & Dvorak, D. R. *J. Insect Physiol.* **29**, 547–560 (1983).
- Egelhaaf, M., Borst, A., Warzecha, A. K., Flecks, S. & Wildemann, A. *J. Neurophysiol.* **69**, 340–351 (1993).
- Hausen, K. *Biol. Cybern.* **45**, 143–156 (1982).
- Hausen, K. *Biol. Cybern.* **46**, 67–79 (1982).
- Haag, J. thesis, Univ. Tuebingen (1994).
- Borst, A., Egelhaaf, M. & Haag, J. *J. comput. Neurosci.* **2**, 5–18 (1995).
- Reichardt, W. *J. comp. Physiol.* **A161**, 533–547 (1987).
- Borst, A. & Egelhaaf, M. *Trends Neurosci.* **12**, 297–306 (1989).
- Borst, A. & Egelhaaf, M. in *Visual Motion and its Role in the Stabilization of Gaze* (eds Miles, F. A. & Wallman, J.) 3–27 (Elsevier, Amsterdam, 1993).
- Gauck, V., Egelhaaf, M. & Borst, A. in *Göttingen Neurobiology Report 1994* (eds Elsner, N. & Breer, H.) 445 (Thieme, Stuttgart, 1994).

ACKNOWLEDGEMENTS. We are grateful to T. Brotz and J. Gold for carefully reading the manuscript.

The rate of information transfer at graded-potential synapses

R. R. de Ruyter van Steveninck*† & S. B. Laughlin†

* NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

† Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

THERE are few estimates of the rates at which spiking neurons transmit information¹⁻⁵, and none for synapses transmitting graded signals. We have measured the rates at which blowfly (*Calliphora vicina*) photoreceptors transmit information through chemical synapses to large monopolar cells (LMCs). The graded responses of these non-spiking cells transmit as much as 1,650 bits per second, five times the highest rates measured in spiking neurons⁶. The widespread occurrence of non-spiking neurons in sensory systems could well reflect this superior performance. Comparing transmission rates in pre- and post-synaptic cells, we estimate that each synaptic active zone transmits ~ 50 bits s^{-1} . This estimate assumes statistical independence of the noise generated at active zones and makes use of a detailed morphometric analysis of photoreceptor-LMC synapses^{6,7}. These measurements provide a benchmark for quantifying the performance of synaptic mechanisms and for understanding the limitations that synaptic transmission and spike coding place upon neural computation.

Information capacity is a basic measure, specifying the logarithm of the number of distinct messages that a signalling system can transmit per unit time. To estimate the information capacities of neurons, we must measure the reliability and dynamics of the neural code¹⁻⁵. Photoreceptors of the fly compound eye code small intensity fluctuations as analogue changes in membrane potential. These graded potentials drive synapses that generate graded responses in the second-order neurons, the LMCs. For the functionally appropriate signals used here, signal and noise are approximately gaussian. For these conditions, Shannon⁸ derived R , the rate of information transmission in bits s^{-1} , as

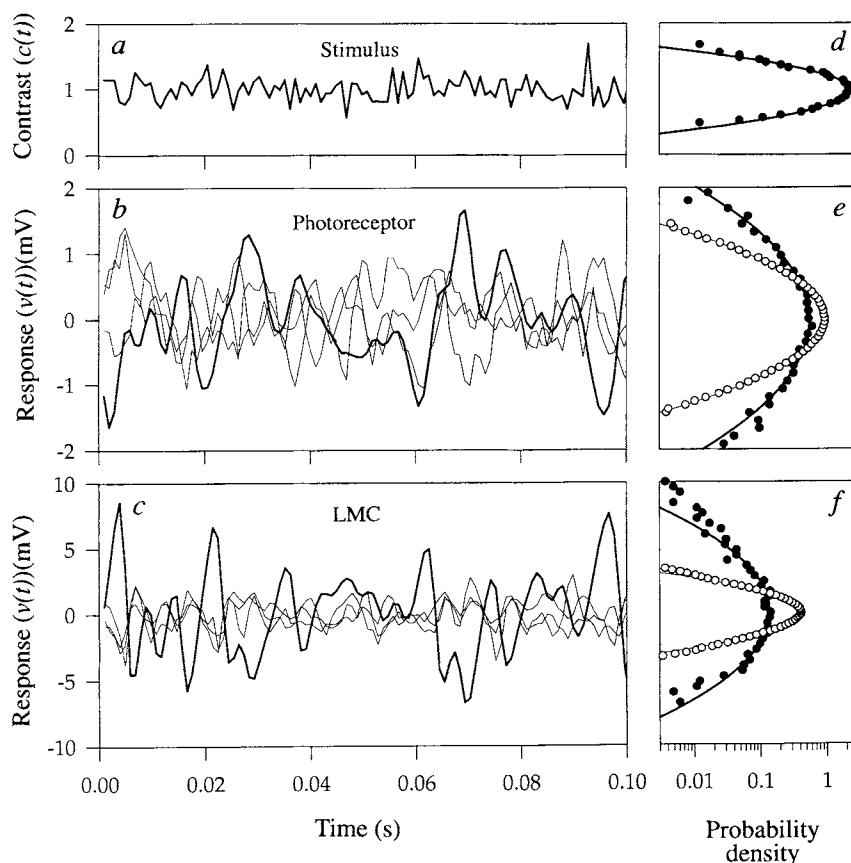
$$R = \int_0^\infty df \log_2 \left[1 + \frac{S(f)}{N(f)} \right] \quad (1)$$

where $S(f)$ and $N(f)$ are the power spectral densities of signal and noise respectively. Note that this formulation considers both reliability and dynamics by treating the signal-to-noise ratio in the frequency domain.

Coding by photoreceptors and LMCs is approximately linear⁹⁻¹¹, enabling us to measure signal and noise spectral densities by recording graded-potential responses to a pseudorandom sequence of light intensity fluctuations. Here we specify this stimulus by the contrast modulation $c(t)$ (Fig. 1). The encoded signal is the average response, $v_{av}(t)$, to repeated presentations of $c(t)$. Note that stimulus and response are continuous functions of time, not single numbers. Each response to a single presentation deviates from the average, and from these random deviations we obtain the noise power spectral density $N(f)$ (Fig. 1).

Because coding is linear, it is described by a transfer function $T(f)$, which here is the contrast to voltage gain at each frequency.

FIG. 1 Examples of raw data traces of the stimulus contrast signal (a), and the average voltages as well as the fluctuations around the average recorded from a photoreceptor (b) and an LMC (c). d-f, Probability distributions corresponding to these signals. Intracellular recordings were made from photoreceptors and LMCs in the retinas of intact blowflies, *Calliphora vicina*, using conventional microelectrode recording techniques¹⁹. All experiments were done at room temperature, 22–24 °C. The stimulus source was a high-intensity green LED modulated by a 2-s pseudorandom contrast sequence. Here we define contrast as $c(t) = I(t)/I_0$, with $I(t)$ the light intensity as a function of time, and I_0 its time-average; thus $c(t)$ is dimensionless with a time average of 1. This sequence, a short segment of which is shown in a, was repeated 76 times while an on-line computer recorded the set of cell voltage responses $\{v_i(t)\}$. The ensemble average of this set, $v_{av}(t)$ (thick lines in b and c), approximates the cell's noiseless response to $c(t)$. Subtracting $v_{av}(t)$ from each of the 76 voltage records gave the set of noise traces $[v_i(t) - v_{av}(t)]$, and averaging the power spectra of these traces gave the overall noise power density spectrum $N(f)$. Three representative noise traces are shown by the thin lines in b and c. In most cases, as in this example, the modulated stimulus $c(t)$ had a gaussian distribution (data points and gaussian fit are shown in d) with a flat contrast power density of $\sim 3.1 \times 10^{-5} \text{ Hz}^{-1}$ up to 500 Hz. As expected⁹⁻¹¹, photoreceptors responded linearly to this stimulus, producing a gaussian distribution of signal levels (filled circles in e) and of noise (open circles in e). Gaussian fits to the data points are shown as the continuous lines. LMC linearity was checked explicitly by recording responses to two different contrast waveforms and to their superposition. The response to the summed stimulus was very close to the sum of the responses to the individual random waveforms, as required for linearity. The distributions of LMC signal (filled circles in f) and noise (open circles in f) deviated somewhat from gaussian (continuous lines). However, for computing the information rate, the approximation is still quite reasonable. This was checked by comparing



the entropy difference between the measured signal and noise distributions (2.81 bits) to the entropy difference between the fitted signal and noise distributions (3.03 bits). With these conditions, equation (1) gives a good approximation to the information transmission of the cells.

$T(f)$ is determined directly as the ratio of the frequency components of the measured average voltage signal and the stimulus contrast: $T(f) = V_{av}(f)/S_c(f)$, where $V_{av}(f)$ and $S_c(f)$ are the Fourier transforms of $v_{av}(t)$ and $c(t)$. To combine data from different cells, we use a common measure appropriate for both

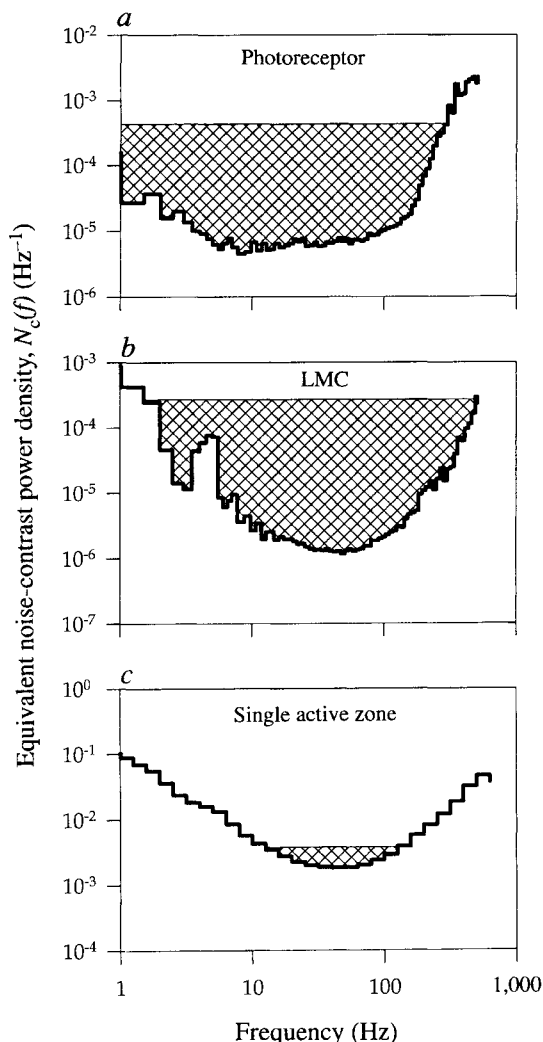


FIG. 2 An account of the procedure used to measure information transfer rates and information capacities. Equivalent contrast noise spectral densities (thick lines), and distribution of optimal stimulus contrast spectral density (cross-hatched areas) for a photoreceptor (a), an LMC (b), and a single synaptic active zone (c). Note the differences in vertical scales, and in particular that the total surface area in the cross-hatched areas in all three panels is equal to 0.1. Shannon⁸ considered the problem of optimal signal transmission through a gaussian channel, given that the total transmitted power is limited. Our problem is the same, except that we fix the total signal power, that is, we fix the stimulus contrast spectral density integrated over frequency. The solution⁸ of the optimization problem is to distribute the contrast power over the different frequency bands in such a way that the summed contribution of signal plus noise power densities is flat where that is possible. This is analogous to filling a vessel shaped as the equivalent contrast noise spectrum with an amount of water that is representative of the total stimulus contrast power. This filling was performed by an iterative procedure, with a fixed total contrast power of 0.1. From the resulting contrast power spectral density, $S_c(f)$, we then derive the information capacity by equation (1). The equivalent contrast power spectral density shown in c is based on interpolated experiment data (Fig. 4). This particular example was computed for the same bump rate as the data shown in b. Note that here we compute the information capacity for the transfer of contrast information. This is not quite the same as the information capacity for transfer of pre- to postsynaptic voltage. The error is very small, as the spectra are quite flat in the region of interest; moreover, the error will lead to an underestimate of the true value.

photoreceptors and LMCs, namely the power spectral density of equivalent contrast noise, $N_c(f)$ (Fig. 2). This is defined as a stimulus contrast power spectral density that, after filtering by $T(f)$, would have the same spectral density as the measured noise: $N_c(f) = N(f)/|T(f)|^2$.

The information transmission rate R now follows from substituting $N_c(f)$ and $S_c(f)$ into equation (1). However, R depends not only upon system performance, as measured by $N_c(f)$, but upon one's choice of signal, $S_c(f)$. A more universal measure is the information capacity, R_{max} , defined as the maximum transmission rate, achieved by selecting the stimulus $S_c(f)$ that optimally drives the cell. Because R_{max} is increased by boosting stimulus contrast, the contrast variance of $S_c(f)$ was fixed at 0.1. This is a natural choice, as it is representative of natural levels^{12,13} and falls within the linear response range of photoreceptors and LMCs. Having fixed the total contrast, we derive $S_c(f)$ for the optimum stimulus (Fig. 2), and use this to calculate the information capacity R_{max} (equation (1)).

Data from five photoreceptors and three LMCs produce consistent estimates of information capacity (Fig. 3). Information capacities rise steadily with mean intensity, as expected of a system partly limited by photon shot noise. LMCs demonstrate the benefits of the convergence of signals from six photoreceptors¹⁴⁻¹⁶, achieving a capacity of 1,650 bits s⁻¹, as compared with 1,000 bits s⁻¹ in photoreceptors. Note that, because of the redundancy inherent to parallel transmission of the same signal, the post-synaptic information capacity is far less than the sum of the presynaptic capacities.

The synaptic information capacity can be deduced by comparing the pre- and postsynaptic equivalent noise spectral densities, $N_{cpre}(f)$ and $N_{cpost}(f)$. The comparison must be between cells

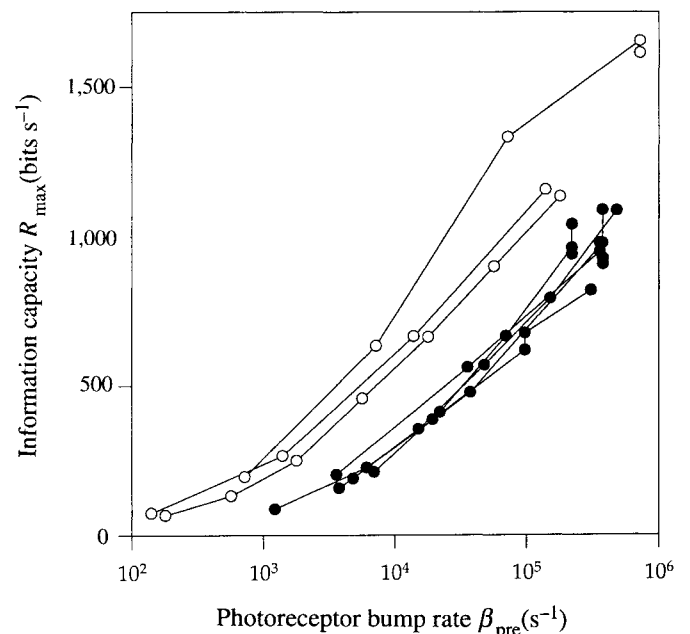


FIG. 3 Information capacities of photoreceptors and LMCs as a function of mean light intensity, computed according to equation (1), with an optimal contrast stimulus as defined in the legend to Fig. 2. Light intensity along the abscissa is expressed as the number of photoconversions per photoreceptor per second. This measure of effective intensity was extrapolated from counts of quantum bumps (electrical events triggered by single photon absorptions¹⁷) at low light levels and the values of the calibrated neutral density filters used to attenuate the light source. The bump rate measured in LMCs was divided by 6 to give the bump rate in each of the 6 pre-synaptic photoreceptors. Lines connect points belonging to an individual cell. The high information rates assume an optimal stimulus but are not unreasonable: with non-optimized but stronger stimuli (total contrast power, 0.28), we directly measured LMC transinformation rates of 1,150 bits s⁻¹.

driven by stimuli of equivalent quantal content, because bit rates depend on photon flux (Fig. 3). Strict equivalence could not be achieved by recording simultaneously from a photoreceptor and its postsynaptic LMC. Instead we calibrated the quantum catch of each recorded cell and corrected for the small differences. Calibration was done by counting the rate of production of

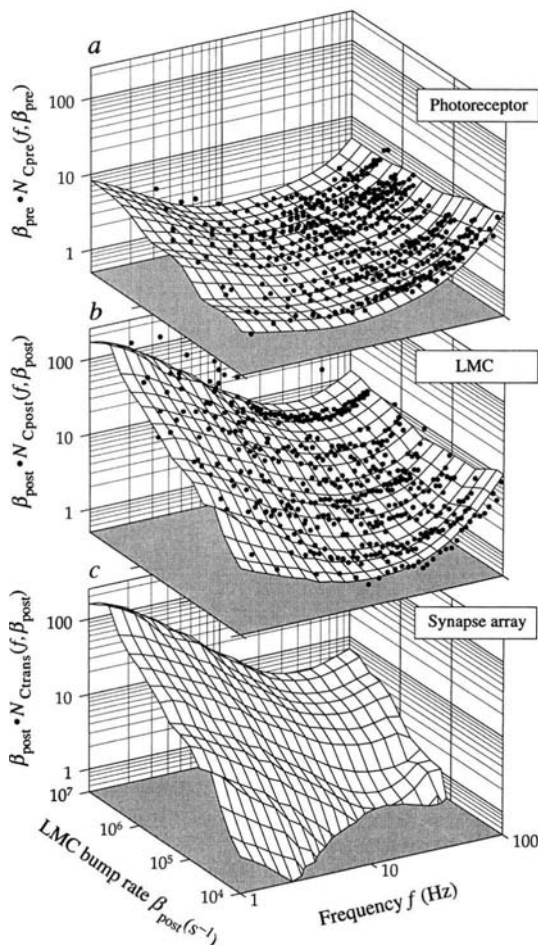


FIG. 4 a, b, Measured values (black dots) of $\beta \cdot N_c(f, \beta)$ for three photoreceptors (a) and three LMCs (b), together with smooth surfaces obtained from interpolation of the data points. The values of β specified on the left-hand axes are the postsynaptic LMC bump rate β_{post} . The presynaptic rate is $\beta_{\text{pre}} = \beta_{\text{post}}/6$. Equivalent contrast power spectral densities for each value of β are computed as $N_c(f, \beta) = N(f, \beta)/|T(f, \beta)|^2$. Shot noise analysis shows that, owing to photon poisson statistics, an ideal photon counter absorbing photons at the extrapolated bump rate should have $\beta \cdot N_c(f, \beta) = 1$. A realistic photon counter then obeys $\beta \cdot N_c(f, \beta) \geq 1$, or in other words, $(\beta \cdot N_c(f, \beta))^{-1}$ is a measure of the detector's quantum efficiency. For the photoreceptors and LMCs studied here, this must be interpreted as a measure of efficiency relative to the dark-adapted state, in which the bump count was calibrated. From about 5 to 50 Hz, the photoreceptor's performance is within a factor of two from the ideal behaviour. At its best frequency, the LMC stays within a factor of 7 from the limit of 10^7 s^{-1} set by the bump rate. For both cells, the efficiency decreases at high frequencies, presumably as a result of time jitter in the transduction process. The decrease in efficiency at low frequencies results at least partly from instrumental noise. The interpolated surfaces give a fair representation of the measured data for each cell type, indicating that $N_c(f, \beta)$ is a good universal description of the behaviour of each cell type. c, The equivalent contrast noise contribution of the full synaptic array. The figure shows $\beta \cdot N_{\text{Ctrans}}(f, \beta)$ (see equation (3)), which is computed by subtracting the surface in a from that in b. $\beta \cdot N_{\text{Ctrans}}(f, \beta)$ must be interpreted as the contribution of synaptic transmission noise relative to the photon shot noise. At low values of β , we see that $\beta \cdot N_{\text{Ctrans}}(f, \beta) \ll 1$, so photon noise dominates the synaptic noise. Values below 0.5 are not shown, as low values are hard to estimate reliably. At higher intensities, $\beta \cdot N_{\text{Ctrans}}(f, \beta)$ increases, indicating that synaptic transmission, rather than the photon flux, becomes the information bottleneck¹⁹.

quantum bumps, discrete voltage pulses produced at very low light levels by single photon hits¹⁷⁻¹⁹. When intensity was increased in calibrated steps, the low-level count was increased by the same factor to give the extrapolated bump rate, β . From measurements of $N_c(f)$ made at different values of β , we interpolate to derive $N_c(f, \beta)$, the function that specifies the equivalent contrast noise spectrum for different β values (Fig. 4).

We calculate synaptic transmission rates as follows. Six photoreceptors carrying identical signals converge upon a single LMC and contribute equally to the postsynaptic response^{6,16}. Thus, when transmitting the same signal, the bump rate in a photoreceptor is one sixth that in an LMC: $\beta_{\text{pre}} = \beta_{\text{post}}/6$. The six photoreceptors contribute independent noise so that convergence improves the signal-to-noise ratio, reducing the equivalent contrast power of presynaptic noise by a factor of 6. Given that the total postsynaptic noise is the sum of noise injected by photoreceptors and noise introduced during transmission, $N_{\text{Ctrans}}(f, \beta)$, we have:

$$N_{\text{Cpost}}(f, \beta_{\text{post}}) = \frac{N_{\text{Cpre}}(f, \beta_{\text{pre}})}{6} + N_{\text{Ctrans}}(f, \beta_{\text{post}}) \quad (2)$$

Calculating $N_{\text{Ctrans}}(f, \beta)$ from equation (2), we obtain a synaptic information capacity of $2,110 \text{ bits s}^{-1}$ at our highest light intensity. This synaptic rate exceeds the LMC rate of $1,650 \text{ bits s}^{-1}$ because it is equivalent to the capacity of an LMC that is driven by noise free photoreceptors.

Photoreceptor-LMC synapses resemble their graded response counterparts in the vertebrate retina in several respects. Their active zones, measuring $0.5 \times 0.2 \mu\text{m}$, contain a high density of vesicles grouped around a prominent presynaptic structure⁷. Each active zone faces a tetrad of postsynaptic elements in which the dendrites of two LMCs form the central pair. In the housefly *Musca* each of the six presynaptic photoreceptors drives an LMC with approximately 200 separate and regularly spaced active zones, giving a total of approximately 1,200 zones driving each LMC⁶. With an optimal synaptic coding strategy, every zone would carry a unique signal component, at a rate of $2,110/1,200 = 1.8 \text{ bits s}^{-1}$. However, this strategy requires sophisticated encoding mechanisms (for example, tuning individual zones to a personalized narrow band of signal frequencies or decorrelation through local spatial interactions) for which there are no precedents in graded synapses. Indeed, the active zones are structurally identical^{6,7}, and appear to act in parallel to induce additive postsynaptic responses, with approximately linear transmission from photoreceptor to LMC^{16,19,20}. Thus the available evidence suggests that this array of active zones is employed to average out synaptic noise²¹, leading to considerable redundancy because each zone carries, on average, the same signal. With m zones in parallel, the synaptic response amplitude and the total synaptic noise power both increase by a factor of m . Consequently, for the full array

$$N_{\text{Ctrans}}(f) = \frac{m \cdot N_{\text{zone}}(f)}{[m \cdot |T_{\text{zone}}(f)|]^2} = \frac{N_{\text{zone}}(f)}{m \cdot |T_{\text{zone}}(f)|^2} = \frac{N_{\text{Czone}}(f)}{m} \quad (3)$$

with $N_{\text{Czone}}(f)$ the equivalent contrast spectral density of a single zone. Computing $N_{\text{Czone}}(f)$ from equations (2) and (3) at the highest β and combining this with an optimized stimulus of contrast variance 0.1 (Fig. 2), gives (equation (1)) an information capacity of 55 bits s^{-1} per active zone. Owing to redundancy, just as with converging photoreceptors, the total information capacity is far below the sum of the individual capacities. The number of zones is unknown for *Calliphora*, so that we have used data from the smaller fly *Musca*. However, any overestimate of active zone capacity will be small because the neurons of the two species are similar in size²². Moreover, the noise in each zone could be non-gaussian, in which case we underestimate the information capacity.

Our study demonstrates that sensory neurons and chemical synapses are remarkably effective. The information capacities of photoreceptors and LMCs are respectively 3 and 5 times higher

than the maximum rates reported by spiking neurons, measured in cricket cercal afferents³. This large difference implies that graded potential neurons are specialized for fast and accurate signalling over short distances^{23,24}, so explaining their widespread use in sensory systems²¹. Moreover, information can easily be lost when graded synaptic inputs are converted into spike trains and this, together with the information capacity of synapses, must constrain neural network design. Information capacities also set constraints upon the underlying molecular mechanisms. For example, the information capacity of a single active zone, about 50 bits s⁻¹, ultimately depends upon the precision of synaptic vesicle discharge. Any description of synaptic transmission that cannot account for this level of performance is incomplete. Thus measures of information capacity provide benchmarks for understanding the performance and design of neural systems at the network, cellular and molecular levels. □

Received 28 July; accepted 30 November 1995.

1. Eckhorn, R. & Pöpel, B. *Biol. Cybern.* **17**, 7–17 (1975).
2. de Ruyter van Steveninck, R. & Bialek, W. *Proc. R. Soc. London* **B234**, 379–414 (1988).
3. Warland, D., Landolfi, M., Miller, J. P. & Bialek, W., in *Neural Systems: Analysis and Modeling*

- (ed. Eckman, F. H.) 327–333 (Kluwer Academic, Boston, 1993).
4. Bialek, W., Rieke, F., de Ruyter van Steveninck, R. & Warland, D. *Science* **252**, 1854–1857 (1991).
5. Kjaer, T., Hertz, J. & Richmond, B. J. *Comp. Neurosci.* **1**, 109–139 (1994).
6. Nicol, D. & Meinertzhagen, I. A. *J. comp. Neurol.* **207**, 29–44 (1982).
7. Fröhlich, A. *J. comp. Neurol.* **241**, 311–326 (1985).
8. Shannon, C. E. *Proceedings of the Institute of Radio Engineers* **37**, 10–21 (1949).
9. Leutscher-Hazellhoff, J. T. *J. Physiol.* **246**, 333–350 (1975).
10. French, A. S. & Järvilehto, M. *J. comp. Physiol.* **126**, 87–96 (1978).
11. Juusola, M., Uusitalo, R. O. & Weckström, M. *J. gen. Physiol.* **105**, 117–147 (1995).
12. Laughlin, S. B. Z. *Naturf.* **366**, 910–912 (1981).
13. Ruderman, D. & Bialek, W. *Phys. Rev. Lett.* **73**, 814–817 (1994).
14. Braitenberg, V. *Expl. Brain Res.* **3**, 271–298 (1967).
15. Kirschfeld, K. *Expl. Brain Res.* **3**, 248–270 (1967).
16. van Hateren, J. H. J. *comp. Physiol.* **A161**, 849–855 (1987).
17. Fuortes, M. G. F. & Yeadle, S. J. *gen. Physiol.* **47**, 443–463 (1964).
18. Howard, J., Blakeslee, B. & Laughlin, S. B. *Proc. R. Soc. Lond.* **B 231**, 415–435 (1987).
19. Laughlin, S. B., Howard, J. & Blakeslee, B. *Proc. R. Soc. Lond.* **B 231**, 437–467 (1987).
20. van Hateren, J. H. & Laughlin, S. B. *J. comp. Physiol.* **A166**, 437–448 (1990).
21. Laughlin, S. B. *J. comp. Physiol.* **84**, 335–355 (1973).
22. Strausfeld, N. J. *Atlas of an Insect Brain* (Springer, Berlin, 1976).
23. Pearson, K. G. in *Simpler Networks and Behaviour* (ed. Fentress, J. C.) 99–110 (Sinauer, Sunderland, 1976).
24. Bush, B. M. H. in *Neurons without Impulses* (eds Roberts, A. & Bush, B. M. H.) (Cambridge University Press, Cambridge, 1981).

ACKNOWLEDGEMENTS. We thank W. Bialek for suggestions, and I. Meinertzhagen, H. Barlow, D. Ruderman, D. O'Carroll, J. Anderson, P. Newland, D. Bray and R. Hardie for critical comments. R.d.R.v.S. was supported by a Royal Society Science Exchange Fellowship and S.B.L. by grants from the SERC Invertebrate Neuroscience Initiative, BBSRC & EPSRC Image Interpretation Initiative.

Inhibition of acute lymphoblastic leukaemia by a Jak-2 inhibitor

Naftaly Meydan*, **Tom Grunberger***,
Harjit Dadi*, **Michal Shahar***, **Enrico Arpaia***,
Zvi Lapidot†, **J. Steven Leeder***, **Melvin Freedman***,
Amos Cohen*, **Aviv Gazit‡**, **Alexander Levitzki‡**
& **Chaim M. Roifman*§**

* Divisions of Immunology/Allergy, Haematology and Pharmacology, The Hospital for Sick Children, and the Department of Pediatrics, University of Toronto, Toronto M5G 1X8, Canada

† Department of Immunology, The Weizmann Institute, Rehovot, Israel

‡ Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University, Jerusalem, Israel

ACUTE lymphoblastic leukaemia (ALL) is the most common cancer of childhood. Despite the progress achieved in its treatment, 20% of cases relapse and no longer respond to chemotherapy. The most common phenotype of ALL cells share surface antigens with very early precursors of B cells and are therefore believed to originate from this lineage^{1,3}. Characterization of the growth requirement of ALL cells indicated that they were dependent on various cytokines, suggesting paracrine and/or autocrine growth regulation^{4–6}. Because many cytokines induce tyrosine phosphorylation in lymphoid progenitor cells, and constitutive tyrosine phosphorylation is commonly observed in B-lineage leukaemias^{7,8}, attempts have been made to develop protein tyrosine kinase (PTK) blockers of leukaemia cell growth^{9,10}. Here we show that leukaemic cells from patients in relapse have constitutively activated Jak-2 PTK. Inhibition of Jak-2 activity by a specific tyrosine kinase blocker, AG-490, selectively blocks leukaemic cell growth *in vitro* and *in vivo* by inducing programmed cell death, with no deleterious effect on normal haematopoiesis.

We analysed a large selection of tyrphostins as potential blockers of the growth of relapsed ALL cells. Among the tyrphostins tested (Fig. 1a), AG-490 was the most effective inhibitor of DNA synthesis in multiple ALL cell lines. The dose-dependent inhibition of DNA synthesis in G₂, A₁ and C₁ leukaemia cell lines is

shown in Fig. 1b. At 5 μM AG-490, growth of all pre-B ALL cells was almost completely blocked. Other tyrphostins (Fig. 1a), such as the epidermal growth factor-receptor inhibitor quinazoline AG-1478 (ref. 11), AG-18 and AG-1007, HER-2/neu kinase inhibitor AG-879, and the B-cell antigen-receptor blockers AG-30 and AG-126 (ref. 12), had no inhibitory effect on DNA synthesis in ALL cells. Furthermore, AG-490 had no significant effect on the growth of mitogen-stimulated normal B or T cells, B-cell lymphoma (Ramos) or T-cell leukaemia (Molt-13) cells (Fig. 1b). The stringent structural requirement for growth inhibition by tyrphostins suggests that AG-490 has a specific effect on leukaemic cells as a result of inhibition of tyrosine kinase activity rather than nonspecific toxicity. Support for this hypothesis comes from the dose-dependent inhibition of tyrosine phosphorylation by AG-490 in all ALL cells tested (Fig. 2a). These results indicate that the tyrphostin AG-490 specifically inhibits the activity of a protein tyrosine kinase that is constitutively activated in all of the leukaemic cells tested.

Because many lymphokine receptors share common PTK signalling pathways^{13–19}, we examined the possibility that one (or more) PTK could be linked to the uncontrolled growth of leukaemic cells. We therefore systematically assessed the identity of the cytoplasmic PTKs in leukaemic cells and their sensitivity to AG-490. We first examined the Jak family of kinases, which are associated with growth-factor receptors in haematopoietic and lymphoid cells^{13–18}. This family has four members Jak-1, Jak-2, Jak-3 and Tyk-2, all of which possess two non-identical kinase domains and unique amino-terminal sequences^{20–22}. Immunoprecipitation and subsequent blotting with antibodies raised against Jak-1, Jak-3 (Fig. 2b) and Jak-2 (Fig. 2c) shows that, in G₂ ALL, Jak-2 is abundantly expressed but Jak-1 and Jak-3 are below detection as compared with activated T cells. Furthermore, Jak-2 is tyrosine-phosphorylated in G₂ cells, (Fig. 2d) but not in normal pre-B cells (Fig. 2e), and its phosphorylation is inhibited by AG-490 in a dose-dependent manner, suggesting a constitutive activation in this kinase in ALL. Consistent with these results, *in vitro* kinase activity of Jak-2 was inhibited in a dose-dependent manner, whether AG-490 was added before (Fig. 2f, g) or during the reaction (data not shown). The AG-490 concentrations that block G₂ cell proliferation and Jak-2 activity *in vitro* (Fig. 2f) and in intact cells (Fig. 2d) are almost identical. *In vitro* kinase activities of Lck, Lyn, Btk, Syk, and Src, which are abundantly expressed in these cells^{23,24}, remained unaffected by AG-490 treatment (Fig. 2h). Identical results were obtained by treating whole cells with AG-490, or adding it directly to the kinase

§ To whom correspondence should be addressed.

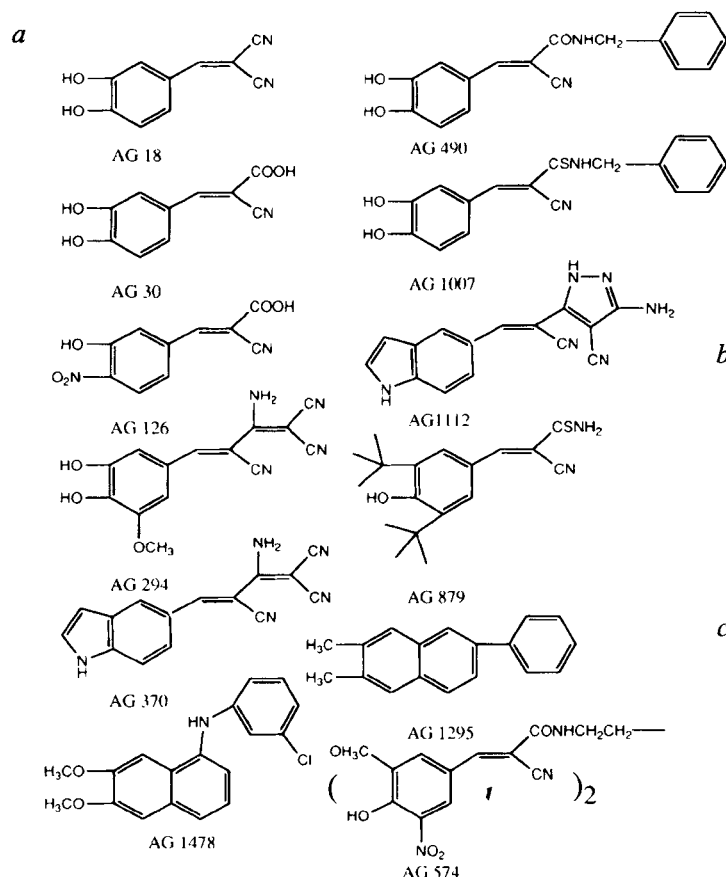
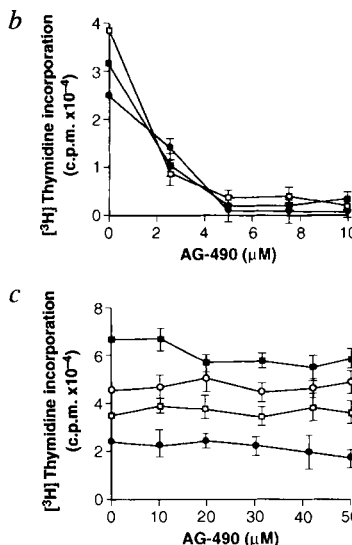


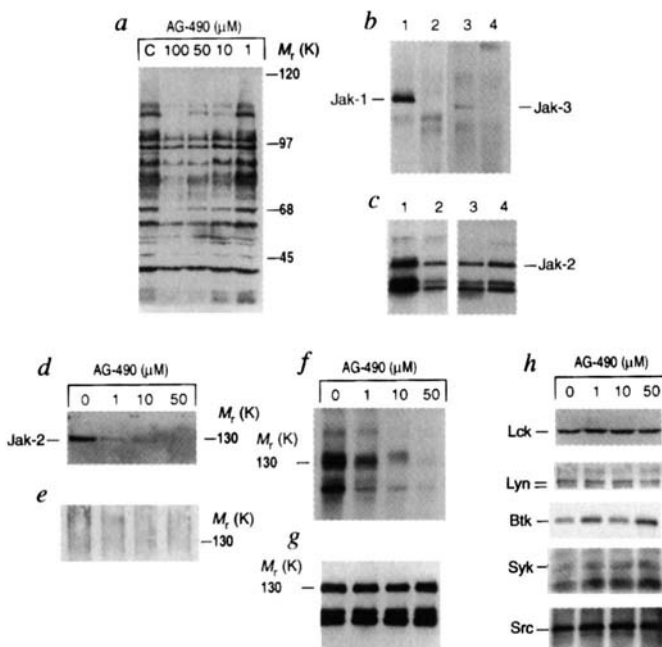
FIG. 1 Screening of tyrphostins that have an antiproliferative effect on leukaemic cells but not on normal cells. **a**, Tyrphostins from different families were tested to determine whether they inhibited pre-B ALL cell growth. Among these tyrphostins, only AG-490 was found to be active. **b**, The effect of AG-490 on [3 H] thymidine incorporation in three different pre-B ALL (CD19⁺, HLA-DR⁺, CD10⁺, Cμ⁻) cell lines derived from patients with relapse leukaemia, including G₂ (filled circles), A₁ (open squares) and C₁ (filled squares). **c**, DNA synthesis in PHA (20 μg ml⁻¹)-stimulated normal T cells (filled squares), SAC-stimulated B cells (filled circles), B lymphoma Ramos cell line (open circles), and Molt-13 T lymphoma cell line (open squares) treated for 16 h with increasing concentrations of AG-490.



METHODS. Cells were incubated at 37 °C in 5% CO₂ with complete culture medium RPMI 1640 supplemented with 10% (v/v) FCS (GIBCO); 2 × 10⁵ cells were seeded in flat-bottomed tissue-culture plates in the presence or absence of different concentrations of AG-490 as indicated. [3 H] thymidine (1 μCi) was added to each well 6 hours or more before the cultures were terminated. Cells were then collected and samples counted in a liquid scintillation counter as previously described^{8,12,19}.

FIG. 2 Effect of AG-490 on protein tyrosine activity in G₂ leukaemic cells. **a**, Inhibition of tyrosine phosphorylation in G₂ cells by AG-490. G₂ leukaemic cells were incubated for 16 h in the absence (lane C) or presence of AG-490 at the concentrations indicated. Phosphotyrosine immunoblots were prepared using standard methods^{12,19}. **b**, Expression of other Jak family members in lymphocytes. Expression of Jak-1 and Jak-3 kinases was determined in activated normal T-cells (lanes 1 and 3) and leukaemic G₂ cells (lanes 2 and 4) by immunoprecipitating Jak-1 or Jak-3 and subsequent immunoblotting with the corresponding antibody. There was no expression of Jak-1 and only traces of Jak-3 in G₂ cells. **c**, Quantitative assessment of Jak-2 in leukaemia cells (lane 1) as compared with normal T cells (lane 2), B cells (lane 3) and fibroblasts (lane 4). Identical amounts of protein obtained from each cell type were loaded. Jak-2 was identified by immunoprecipitation and subsequent immunoblotting with anti-Jak-2 antibody. **d**, Constitutive phosphorylation of Jak-2 is inhibited by AG-490. G₂ leukaemic cells were immunoprecipitated using anti-Jak-2 antibody. Immune complexes were then immunoblotted with anti-phosphotyrosine antibody. **e**, Lack of Jak-2 phosphorylation in unstimulated normal pre-B cells derived from bone marrow. Pre-B cells (CD10⁺ ± 80%) were enriched by sorting CD19⁻ CD10⁺ double-positive cells on a ELIT coulter; analysis was as in **d**. **f**, *In vitro* tyrosine autophosphorylation of Jak-2. Jak-2 was immunoprecipitated from lysates of G₂ cells pretreated for 16 h with increasing concentrations of AG-490 (0–50 μM). A dose-dependent inhibition of *in vitro* kinase activity was demonstrated by assessing Jak-2 autophosphorylation. **g**, Quantitative assessment of Jak-2 in G₂ cells. To ensure that the reduction in Jak-2 kinase activity is not due to decreased amounts of Jak-2 protein, the same blot (**a**) was reprobed with antibodies raised against Jak-2. **h**, *In vitro* kinase activity of Src family kinases in G₂ cells. Lck, Lyn, Btk, Syk and Src proteins were immunoprecipitated from lysates of G₂ cells with corresponding antibodies, and *in vitro* auto-kinase activity was examined in the presence or absence of various concentrations of AG-490.

METHODS. Immunoprecipitates were prepared as previously described³⁰. Western blotting with anti-phosphotyrosine, anti-Jak-1, anti-Jak-2, or anti-Jak-3 antibody was performed as described in Fig. 2. For *in vitro* kinase assays, immunoprecipitates of Jak-2, Lck, Fyn, Btk, Syk or Src prepared as



described above were washed twice with kinase buffer and resuspended in 25 μl buffer containing varying concentrations of AG-490. The reaction was stopped by the addition of 50 μl SDS-PAGE reducing sample buffer and separated on an 8% SDS-PAGE gel. The products were transferred onto nitrocellulose membrane and visualized by autoradiography.

reaction (not shown). These results strongly support the hypothesis that Jak-2 is the specific target for AG-490 action in ALL cells.

To assess the effect of AG-490 on the growth of freshly obtained primary ALL cells, we have grown ALL cells in methylcellulose using leukaemic blast colony assays²⁵. A concentration-dependent inhibition of colony formation by AG-490 is observed where at 2–5 μ M AG-490, almost complete inhibition of primary or relapsed ALL cell growth is achieved (Table 1a). Remarkably, AG-490 up to a concentration of 20 μ M did not alter growth and maturation of normal haematopoietic progenitor cells, as assessed by the colony-forming assay (Table 1a). This highlights the selective effect of this compound on leukaemia cells.

ALL cells treated with AG-490 were inspected under light microscopy, and morphological changes were observed, over 5 days in culture, that are characteristic of cells undergoing apoptosis. By using flow cytometric analysis of subdiploid DNA²⁶, we show that G₂ cells incubated for 12, 48 and 72 h in the presence of 5 μ M AG-490 displayed increases of 11, 18 and 25% in apoptotic

cells, respectively, as compared to untreated cells (data not shown). Consistent with recent results²⁷, our data raise the possibility that blocking PTK activity might mimic apoptosis induced by deprivation of growth factor.

We have used the SCID mouse model^{28,29} to determine the efficacy of AG-490 against ALL cells *in vivo*. Intravenous injection of ALL cells into SCID mice results within two weeks in tumour proliferation and dissemination into the bone marrow and other organs in a fashion analogous to the terminal stages of ALL in children²⁹. Using a similar protocol, SCID mice were transplanted with G₂ cells and were assigned to receive either no treatment, AG-490 or DMSO delivered by continuous pump infusion supplemented with daily intraperitoneal injections (Table 1b). In some experiments treatment was administered immediately after transplantation (early treatment), whereas in others AG-490 treatment was delayed until four weeks after transplantation to mimic leukaemia presentation in humans. As shown in Table 1b, 4–6 weeks after G₂ transplantation, the bone marrow of

TABLE 1 Detection of leukaemia and the effect of AG-490

(a) Effect of AG-490 on leukaemia formation								
Culture conditions	Normal bone marrow			Leukaemia cell lines (colonies per culture)		Fresh leukaemia (colonies per culture)		
	BFU-E	CFU-GEMM	CFU-GM	G ₂	C ₁	P1	P2	P3
AG-490 (μ M) 20	186 $\pm$ 11	21 $\pm$ 3	744 $\pm$ 116	0	0	0	0	0
15	ND	ND	ND	0	0	1 $\pm$ 1	0	0
10	300 $\pm$ 35	12 $\pm$ 2	360 $\pm$ 22	0	0	1 $\pm$ 1	0	0
5	335 $\pm$ 17	43 $\pm$ 5	544 $\pm$ 40	0	0	2 $\pm$ 1	9 $\pm$ 2	1 $\pm$ 1
2.5	ND	ND	ND	0	7 $\pm$ 2	7 $\pm$ 3	9 $\pm$ 3	5 $\pm$ 1
1	ND	ND	ND	2 $\pm$ 1	5 $\pm$ 1	10 $\pm$ 2	24 $\pm$ 7	7 $\pm$ 1
0.1	ND	ND	ND	38 $\pm$ 4	65 $\pm$ 9	22 $\pm$ 4	75 $\pm$ 11	29 $\pm$ 4
0.01	ND	ND	ND	205 $\pm$ 34	242 $\pm$ 12	30 $\pm$ 6	130 $\pm$ 25	61 $\pm$ 5
Medium	250 $\pm$ 44	12 $\pm$ 3	302 $\pm$ 28	650 $\pm$ 67	580 $\pm$ 101	33 $\pm$ 9	195 $\pm$ 33	72 $\pm$ 10
DMSO 0.00003– 0.05%	225 $\pm$ 19	13 $\pm$ 2	310 $\pm$ 41	610 $\pm$ 115	645 $\pm$ 82	24 $\pm$ 3	210 $\pm$ 36	60 $\pm$ 9
AG-574 20 μ M	466 $\pm$ 50	37 $\pm$ 6	440 $\pm$ 95	798 $\pm$ 80	ND	ND	ND	ND
(b) Detection of human leukaemia in SCID mice								
Weeks after transplant	Treatment (no. of mice)	Plasma AG-490 (μ M) (no. of mice)	CD45 ⁺ cells(%)		HLA-DR ⁺			
			Bone marrow	Spleen	Bone marrow	Spleen		
4–6	None (15)	UD	48 $\pm$ 11	38 $\pm$ 15	46 $\pm$ 3	22 $\pm$ 4		
	DMSO (4)	UD	41 $\pm$ 9	39 $\pm$ 5	51 $\pm$ 17	17 $\pm$ 5		
	AG-490 (12)	<1 (2 [§] , 2 [†])*	18 $\pm$ 6	15 $\pm$ 10	12 $\pm$ 3	7 $\pm$ 4		
	AG-490	1–3 (4 [§] , 1 [†])†	7 $\pm$ 2	6 $\pm$ 4	3 $\pm$ 1	2 $\pm$ 1		
	AG-490	>3 (4 [§] , 2 [†])‡	<1	<1	<1	<1		
7–9	None (15)	UD	74 $\pm$ 17	62 $\pm$ 21	97 $\pm$ 23	45 $\pm$ 19		
	DMSO (3)	UD	68 $\pm$ 4	63 $\pm$ 24	95 $\pm$ 11	54 $\pm$ 22		
	AG-490 (13)	<1 (5 [§] , 2 [†])§	26 $\pm$ 11	21 $\pm$ 15	18 $\pm$ 11	10 $\pm$ 3		
	AG-490	1–3 (2 [§] , 1 [†])	12 $\pm$ 9	10 $\pm$ 7	6 $\pm$ 3	4 $\pm$ 1		
	AG-490	>3 (2 [§] , 1 [†])¶	<1	<1	<1	<1		

a, Values shown are the means of duplicate cultures and are representative of four similar experiments. Normal haematopoietic progenitor cells derived from bone marrow and leukaemic cells were grown in a blast colony assay. Leukaemic cells included the pre-B cell lines G₂ and C₁ that were derived from patients with relapse ALL. In addition, freshly obtained leukaemic cells were taken from bone marrow of 3 different patients (P1–P3) with relapse ALL. The cell-marker phenotypes of all leukaemic cells and cell lines tested here are typical of pro-B ALL expressing of CD5⁺, HLA-DR⁺, CD19⁺, C_μ⁺. The effect of AG-574 (Fig. 1), which has no inhibitory activity on ALL, is shown for comparison. The ALL colony assay for freshly obtained leukaemia was as previously described^{5,27}. The resulting cell population was composed of 99% lymphoblasts. Cells (2 $\times$ 10⁵) were then suspended in α -medium supplemented with 10% fetal calf serum (FCS) plus irradiated autologous leukaemic blasts used as feeder cells and methyl cellulose at a final concentration 0.9% (v/v). Leukaemic cell lines were directly plated as described for fresh leukaemia. For evaluation of normal multilineage colony-forming units (CFU), unsorted nucleated bone-marrow cells were cultured in methylcellulose with Iscov's modified Dulbecco's medium, 30% FCS, GM-CSF and IL-3. Cultures were evaluated for the number of erythroid burst-forming units (BFU-E) colonies of erythroid cells, granulocyte-macrophage colony-forming units (CFU-GM), and mixing (CFU-GEMM) colonies of more than two haematopoietic lineages. All cultures were evaluated after 14 days for the number of colonies. ND, not done; DMSO, dimethylsulphoxide; results are the mean $\pm$ s.e. of controls set up in parallel to all tyrphostin concentrations. b, Because the half-life of AG-490 was estimated to be short, mice received daily intraperitoneal injections of 0.5 mg AG-490, as well as intradermal pump that delivered 0.85 mg AG-490 per day; treatment began immediately after or 4 weeks after the injection of G₂ cells. Mice were killed, and AG-490 plasma levels were determined using reverse-phase HPLC separation, performed with a Waters 6000A pump, an Ultrasphere 150 $\times$ 4.5-mm 5 μ ODS column and Shimadzu SPD-6A UV detector at a wavelength of 254 nm. The solvent was acetonitrile-triethylamine-acetic acid-water (40:0.1:1:60, v/v/v/v) at a flow rate of 1.0 ml min⁻¹. Under these conditions the retention time of AG-490 was 4.7 min. Flow cytometry analysis cells derived from bone marrow or spleen of SCID mice engrafted with G₂ cells. Anti-human CD45⁺ antibody and anti-human HLA-DR antibody recognize human but not mouse CD45 or HLA-DR, respectively. All mice treated with AG-490 had reduced numbers of CD45⁺, HLA-DR⁺ cells in bone marrow and spleen compared with control AG-490 untreated animals or DMSO-treated animals as indicated. UD, undetected; δ , delayed treatment; ϵ , early treatment. Mean concentrations of AG-490 were * 0.6, $\dagger$ 1.9, $\ddagger$ 4.1, $\S$ 0.45, $\parallel$ 2.3 and $\P$ 4.4 μ M.

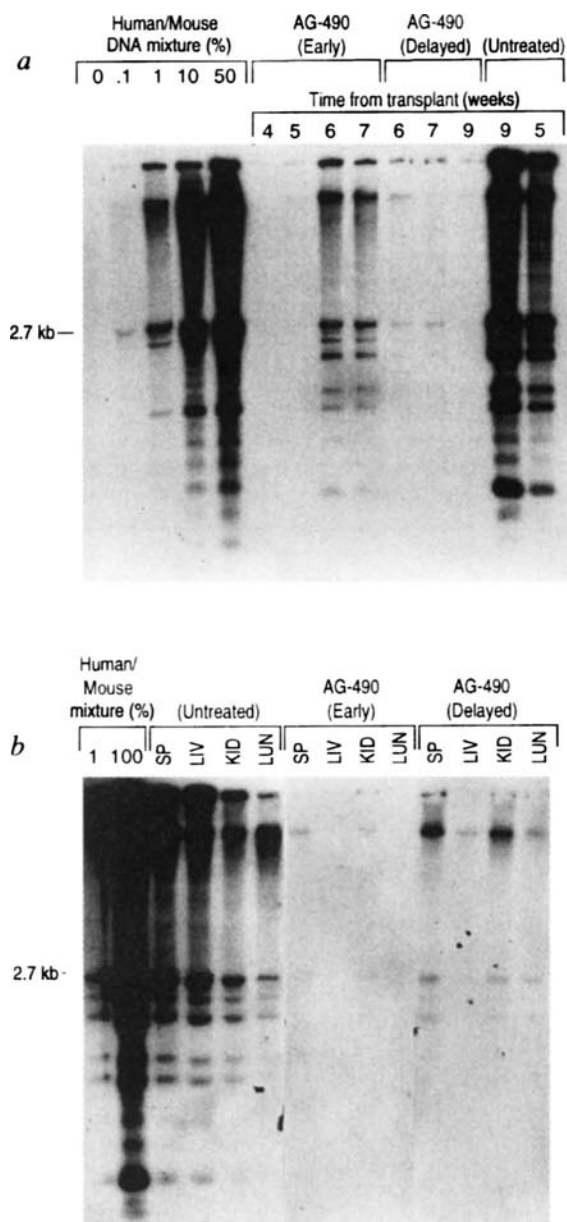


FIG. 3 Analysis of tissues from SCID mice engrafted with G_2 relapse leukaemic cells. G_2 cells (1×10^7) were injected into SCID mice after irradiation with 200 rad. Mice subsequently received AG-490 infusion through a subcutaneous pump (800 μ g per day) and by intraperitoneal injection (500 μ g per day). Treatment began immediately after G_2 injection (early treatment) or 4 weeks after the administration of leukaemic cells (delayed treatment). Each week, some mice were killed, and tissues including bone marrow, spleen (SP), liver (LIV), kidney (KID) and lung (LUN) were obtained for analysis of human DNA. a, The human DNA content in bone-marrow samples obtained from AG-490-treated mice and control engrafted tyrophostin untreated animals as indicated. The percentage of human DNA present in the mouse tissues was estimated using human:mouse DNA mixtures from 0% human:100% mouse to 50% human:50% mouse DNA as indicated. b, The effect of AG-490 in multiple organs of representative mice engrafted with G_2 cells compared to an engrafted but AG-490-untreated mouse. Percentage of human DNA present of human DNA present in the mouse tissues was estimated using human:mouse mixtures as indicated.

METHODS. DNA was phenol-extracted from tissues of transplanted mice using standard techniques. DNA (5 μ g) was digested with *Eco*RI, blotted and probed with p17H8, a human α -satellite probe specific for chromosome 17 sequences^{28,29}. The characteristic 2.7-kb band indicates human sequences.

untreated mice contained 48 and 46%, and the spleen 38 and 22%, of the level of human CD45⁺ and human leukocyte antigen (HLA)-DR⁺ cells, respectively. In sharp contrast, treated mice that had AG-490 plasma levels higher than 3 μ M had virtually undetectable CD45⁺ and HLA-DR⁺ cells in spleen and bone marrow. Other tissues, including lung, kidney, liver and brain, showed an identical pattern (data not shown). This effect of AG-490 was sustained through week 9, despite the increased tumour mass, as shown in Table 1b. Even AG-490 plasma levels less than 3 μ M were still effective in substantially reducing leukaemia infiltrates (Table 1b).

For independent evidence of the effect of AG-490 on ALL cell infiltration in SCID mice, the proportion of human DNA was estimated by hybridization to a human-specific chromosome 17 α -satellite probe using Southern blots^{28,29}. G_2 -engrafted, but not treated, mouse bone-marrow samples contained 20% at 5 weeks and more than 50% at 9 weeks of human DNA levels (2.7-kb band) (Fig. 3). Consistent with the flow cytometry analysis in mice that achieved higher plasma AG-490 levels, only traces of human DNA were found in bone marrow and other organs at 4–9 weeks after transplant (<0.1% to 1%), as shown in Fig. 3a and b, respectively. Furthermore, bone marrow retrieved from these mice was transferred to clonogenic assays and to untreated SCID mice to identify residual leukaemic cells with proliferative potential. Under these experimental conditions, G_2 ALL cells failed to form colonies *in vitro* or engraft into SCID mice that were observed for as long as nine weeks (not shown). These data indicate that the residual human DNA detected in these animals represents human cells incapable of clonal expansion. AG-490 had no deleterious side effects on mouse haematopoiesis, as determined by blood counts and clonogenic assays of murine specimens (not shown).

In summary, we have shown that the PTK inhibitor tyrphostin AG-490 blocks the growth of recurrent pre-B ALL cells *in vitro* and *in vivo* by inducing programmed cell death. This effect of AG-490 is exerted through inhibition of Jak-2 activity that is probably necessary for lymphokine-driven, uncontrolled growth of these leukaemic cells. Improvements in the delivery and pharmacodynamic properties of AG-490 can potentially increase its effectiveness in animal models, and ultimately in the treatment of aggressive relapsed leukaemia. □

Received 9 November; accepted 15 December 1995.

- Nadler, L. M. *et al.* *J. Immun.* **131**, 244–250 (1983).
- Greaves, M. F. *Science* **234**, 697–704 (1986).
- Uckun, F. M. *et al.* *New Engl. J. Med.* **329**, 1296–1301 (1993).
- Williams, G. T., Smith, C. A., Spooncer, E., Dexter, T. M. & Jaylor, D. R. *Nature* **343**, 76–79 (1990).
- Cohen, A. *et al.* *Blood* **78**, 94–102 (1991).
- Touw, J. *et al.* *Blood* **75**, 2097–2101 (1990).
- Grimaldi, J. C. & Meeker, T. C. *Blood* **73**, 2081–2085 (1989).
- Dadi, H., Ke, S. & Roifman, C. M. *Biochem. biophys. Res. Commun.* **192**, 459–464 (1993).
- Burke, T. *et al.* *J. med. Chem.* **36**, 425–432 (1993).
- Cushman, M. *et al.* *J. med. Chem.* **37**, 3353–3362 (1994).
- Oshero, N. & Levitzki, A. *Eur. J. Biochem.* **225**, 1047–1053 (1994).
- Padeh, S., Levitzki, A., Gazit, A., Mills, G. & Roifman, C. M. *J. clin. Invest.* **87**, 1114–1118 (1991).
- Osamu, M. *et al.* *Blood* **84**, 1501–1507 (1994).
- Sato, M. *et al.* *J. exp. Med.* **180**, 2101–2111 (1994).
- Barber, D. L. & D'Andrea, A. D. *Molec. cell. Biol.* **14**, 6506–6514 (1994).
- Tanaka, N. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **91**, 7271–7275 (1994).
- Johnston, J. A. *et al.* *Nature* **370**, 151–153 (1994).
- Wittuhn, B. A. *et al.* *Nature* **370**, 153–157 (1994).
- Roifman, C. M., Wang, G., Freedman, M. & Pan, Z. *J. Immun.* **148**, 1136–1142 (1992).
- Wilks, A. F. *et al.* *Molec. cell. Biol.* **11**, 2057–2065 (1991).
- Harpur, A. G., Andres, A.-C., Zierniecki, A., Aston, R. R. & Wilks A. F. *Oncogene* **7**, 1347–1353 (1992).
- Rothman, P. *et al.* *Immunity* **1**, 457–468 (1994).
- Bolen, J. B., Thompson, P. A., Eiseman, E. & Horak, I. D. *Adv. Cancer Res.* **57**, 103–149 (1991).
- Bolen, J. B. *et al.* *FASEB J.* **6**, 3403–3409 (1992).
- Estrov, Z. & Freedman, M. H. *Exp. Hemat.* **19**, 221–225 (1991).
- Schmid, I., Uittenbogaart, C. H. & Giorgi, J. V. *Cytometry* **15**, 12–20 (1994).
- Uckun, F. M. *et al.* *Science* **267**, 886–891 (1995).
- Kamel-Reid, S. *et al.* *Science* **246**, 1597–1600 (1989).
- Kamel-Reid, S. *et al.* *Leukemia* **6**, 8–17 (1992).
- Arpaia, E., Shahar, M., Dadi, H., Cohen, A. & Roifman, C. M. *Cell* **76**, 947–958 (1994).

ACKNOWLEDGEMENTS. We thank T. Pawson and J. Dick for critical reviews, and S. Wang, D. Petsche, W. Mansvelt and D. Mansvelt for technical assistance. This work was supported by the National Cancer Institute of Canada, the Medical Research Council of Canada, and SUGEN, Inc. N. Meydan was supported by a fellowship from SUGEN, Inc.

Neural correlates of category-specific knowledge

Alex Martin, Cheri L. Wiggs, Leslie G. Ungerleider & James V. Haxby

Laboratory of Psychology and Psychopathology, Building 10, Room 3D-41, National Institute of Mental Health, Bethesda, Maryland 20892, USA

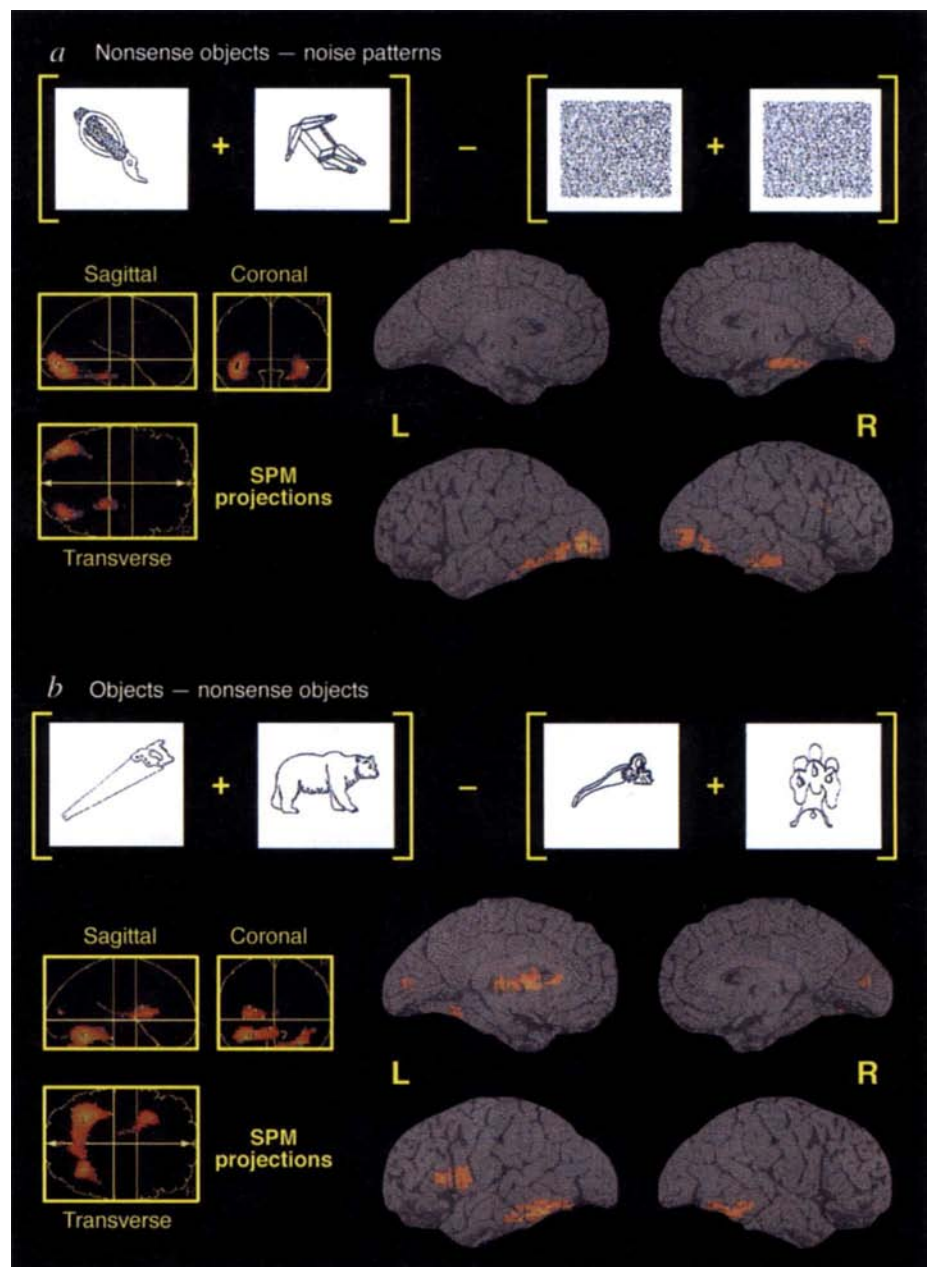
AN intriguing and puzzling consequence of damage to the human brain is selective loss of knowledge about a specific category of objects. One patient may be unable to identify or name living things¹⁻³, whereas another may have selective difficulty identifying man-made objects⁴⁻⁶. To investigate the neural correlates of this remarkable dissociation, we used positron emission tomography to map regions of the normal brain that are associated with naming animals and tools. We found that naming pictures of animals and tools was associated with bilateral activation of the

ventral temporal lobes and Broca's area. In addition, naming animals selectively activated the left medial occipital lobe—a region involved in the earliest stages of visual processing. In contrast, naming tools selectively activated a left premotor area also activated by imagined hand movements⁷, and an area in the left middle temporal gyrus also activated by the generation of action words⁸⁻¹⁰. Thus the brain regions active during object identification are dependent, in part, on the intrinsic properties of the object presented.

We studied sixteen (8 male, 8 female), right-handed subjects with positron emission tomography (PET) to measure changes in regional cerebral blood flow (rCBF) associated with identifying line drawings of animals and of tools¹¹. These categories were chosen because distinctions between four-legged animals are often based on subtle differences in physical features (form, colour and size), whereas distinctions among tools are weighted more heavily towards differences in functional attributes (how objects are used). For each category, subjects named the objects silently during one scan and out loud during another scan to provide behavioural data on naming latency and errors. To

FIG. 1 Regions of increased rCBF ($P < 0.001$) when subjects viewed nonsense objects compared to viewing visual noise patterns (a) were in the left ($-38, -82, -4$) and right ($+34, -72, -8$) inferior occipital/fusiform gyrus, the left temporal fusiform gyrus ($-26, -44, -16$), the right inferior frontal region ($+36, +16, +24$) and the left cerebellum ($-24, -36, -28$) (not shown). Activation was also present in the right parahippocampal gyrus ($+24, -32, -20$; $+28, -6, -28$) and right hippocampus ($+26, -16, -16$). Hippocampal activation was not found when naming real objects was compared to the visual noise baseline, suggesting that the hippocampus may be important in the detection of novelty²². Regions activated ($P < 0.001$) by naming real objects compared to viewing nonsense objects (b) were in the left ($-28, -58, -16$) and right ($+42, -44, -12$) fusiform gyri of the temporal lobes, the left anterior insula/inferior frontal region ($-28, +16, +8$), the left thalamus ($-14, -12, +8$), the calcarine sulcus ($+2, -84, +8$), and the left ($-6, -62, -16$) and right ($+4, -66, -28$) medial cerebellum (not shown). The location of activations are expressed in millimetres as coordinates in the Talairach and Tournoux brain atlas²³.

METHODS. Each stimulus was presented for 180 ms, followed by a centrally located fixation cross for 1,820 ms. Sets of animal and tool drawings were equated for name frequency and category typicality. Different sets were presented during the silent and overt naming conditions, counterbalanced across subjects. PET scans were obtained using a Scanditronix PC2048-15B tomograph (Milwaukee, Wisconsin) which acquires 15 continuous, 6.5-mm-thick cross-sectional images. Within-plane resolution is 6.5 mm (full width at half maximum). Subjects began the task ~30 s before injection of 37.5 mCi of $H_2^{15}O$. Data were analysed using Statistical Parametric Mapping (SPM)²⁴⁻²⁶. Data from each subject were normalized to his/her own global mean flow (ratio correction). Contrasts between tasks were evaluated with *t*-tests, and then converted to *z* scores.



investigate further the object-processing system and to establish baselines for comparison purposes, subjects were also scanned twice while staring at visual noise patterns, and twice while staring at novel nonsense objects¹². The data are shown from the six PET scans during which there was no overt speech (Fig. 1).

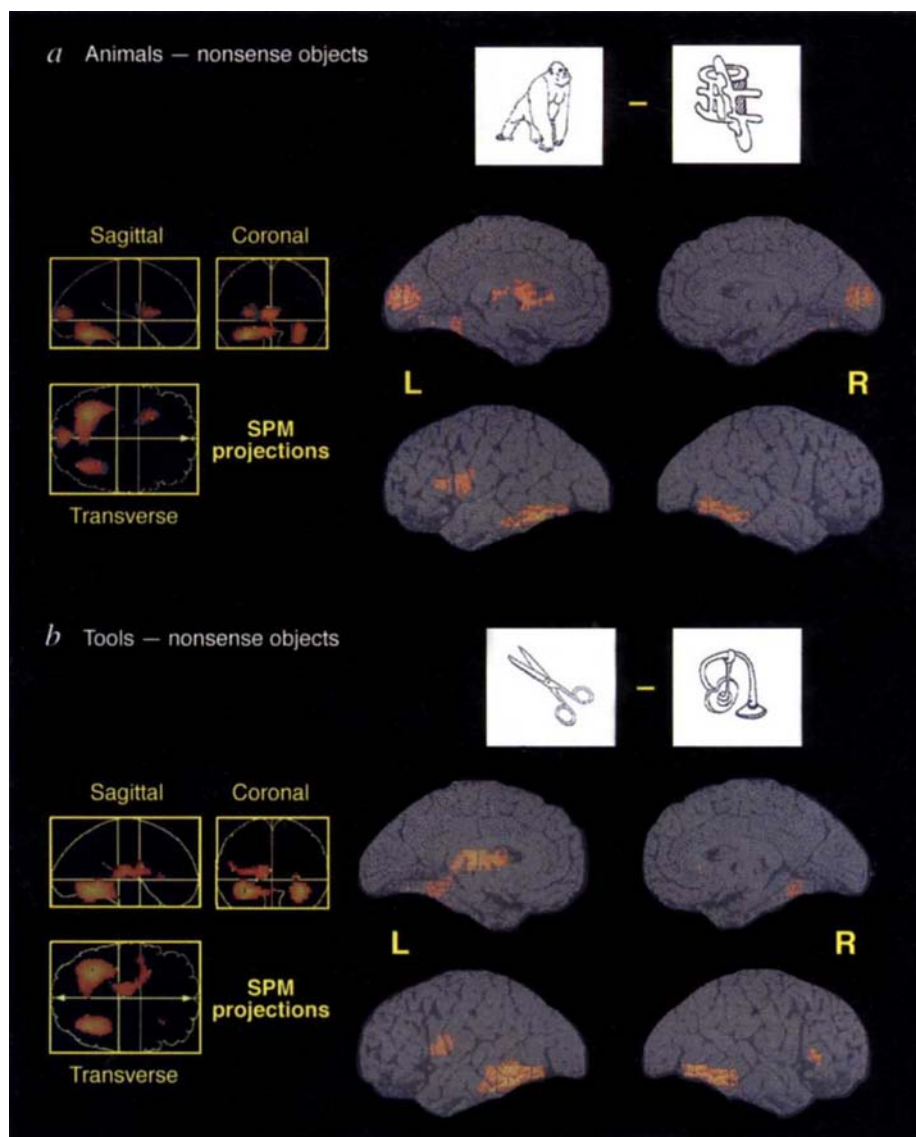
Activation of a hierarchically organized, ventral occipitotemporal object processing stream¹³ was revealed by first comparing perception of noise patterns to nonsense objects (Fig. 1a), and then comparing perception of nonsense objects to silent object naming (animal and tool conditions combined) (Fig. 1b). Relative to perceiving visual noise, perceiving nonsense objects activated the left and right fusiform and inferior gyri of the occipital lobes (Brodmann's area (BA) 19). Relative to viewing nonsense objects, naming real objects also produced bilateral ventral activation that coincided with, and extended anterior to, the activation produced by viewing nonsense patterns. Thus, relative to viewing nonsense objects, the ventral temporal lobes showed increased activity during object naming, whereas the inferior occipital regions did not. Increased rCBF during object naming was also seen in the left insula/Broca's area, as reported in other PET studies of silent word generation⁸ (see figure legends for complete listing of activations).

Identifying real objects also activated the calcarine sulcus (BA 17), or primary visual cortex (Fig. 1b), but calcarine activation occurred only for naming animals (Fig. 2). In addition, although identifying animals and tools were both associated with bilateral

increases in rCBF in the ventral region of the temporal lobes, the left temporal activation extended dorsally to include the middle temporal gyrus for naming tools, but not animals (Fig. 2). Finally, naming tools, but not animals, was associated with activation of the left premotor area (BA 4/6). These category-specific rCBF differences were also observed when the animal and tool conditions were directly contrasted with each other, thus providing strong evidence that rCBF, and hence neuronal activity, was modulated by the type of object presented. Relative to naming tools, naming animals produced a large area of activation in the left medial occipital lobe centred on the calcarine sulcus (Fig. 3), whereas, relative to naming animals, tool-naming produced activity in the left middle temporal gyrus (BA 21) and the left premotor region (BA 6).

Selective activation of the medial aspect of the occipital lobe during animal naming suggests that these stimuli placed greater demands on early stages of visual processing. Indeed, the animal pictures triggered a slower and less accurate response than the tools. However, increased early visual processing demands could not alone account for this finding. The viewing conditions for each task were identical, yet the medial occipital region was activated when naming animals, but not when naming tools, nor when viewing visually complex nonsense objects. To rule out the effects of stimulus complexity, a different group of subjects ($n = 16$) were investigated under the same conditions as in the first study except that objects were presented in silhouette, thereby eliminating

FIG. 2 Regions of increased rCBF ($P < 0.001$) when subjects silently named drawings of animals compared to viewing nonsense objects (a) were in the calcarine sulcus ($-2, -82, +8$), the left ($-24, -58, -16$) and right ($+34, -56, -12$) fusiform gyri of the temporal lobes, the left putamen ($-20, +4, +8$), left thalamus ($-18, -10, +16$), left insula/inferior frontal region ($-28, +14, +8$), and the right lateral ($+32, -60, -20$) and medial ($+8, -56, -28$) cerebellum (not shown). Regions of increased rCBF ($P < 0.001$) when subjects silently named tools compared to viewing nonsense objects (b) were the left ($-28, -50, -12$) and right ($+34, -50, -12$) fusiform gyri of the temporal lobes, left inferior temporal gyrus ($-36, -48, -4$), left putamen ($-24, +4, +8$), left thalamus ($-12, -18, +8$), left insula/inferior frontal region ($-30, +8, +8$), left precentral gyrus (BA 4/6) ($-36, 0, +12$), right inferior frontal gyrus (BA 45) ($+28, +28, +4$), and the right medial cerebellum ($+8, -56, -28$) (not shown). The animal pictures have been judged to be more visually complex and less familiar than the tools¹¹. Consistent with these differences, behavioural data recorded during the overt-naming PET scans indicated that the animal pictures were named more slowly and less accurately than were the tools ($P < 0.01$) (the mean $\pm$ standard error of the mean for voice response time was 620 ± 21.2 ms and error rate was $16.7 \pm 0.77\%$ for naming animals; mean $\pm$ standard error was 566 ± 21.0 ms and the error rate was $10.0 \pm 0.54\%$ for naming tools).



internal detail. This manipulation ruled out differences between the animal and tool stimuli in naming latency and naming errors. Nevertheless, relative to naming tool silhouettes, naming animal silhouettes activated the left medial occipital area (Fig. 3 legend). Selective activation of the occipital cortex during animal naming may reflect top-down modulation, or reactivation, of primary visual areas, rather than increased visual processing during input^{14–16}. Reactivation of this region may be necessary to identify an object uniquely when relatively subtle differences in physical features are the primary means by which the object can be distinguished from other members of its category.

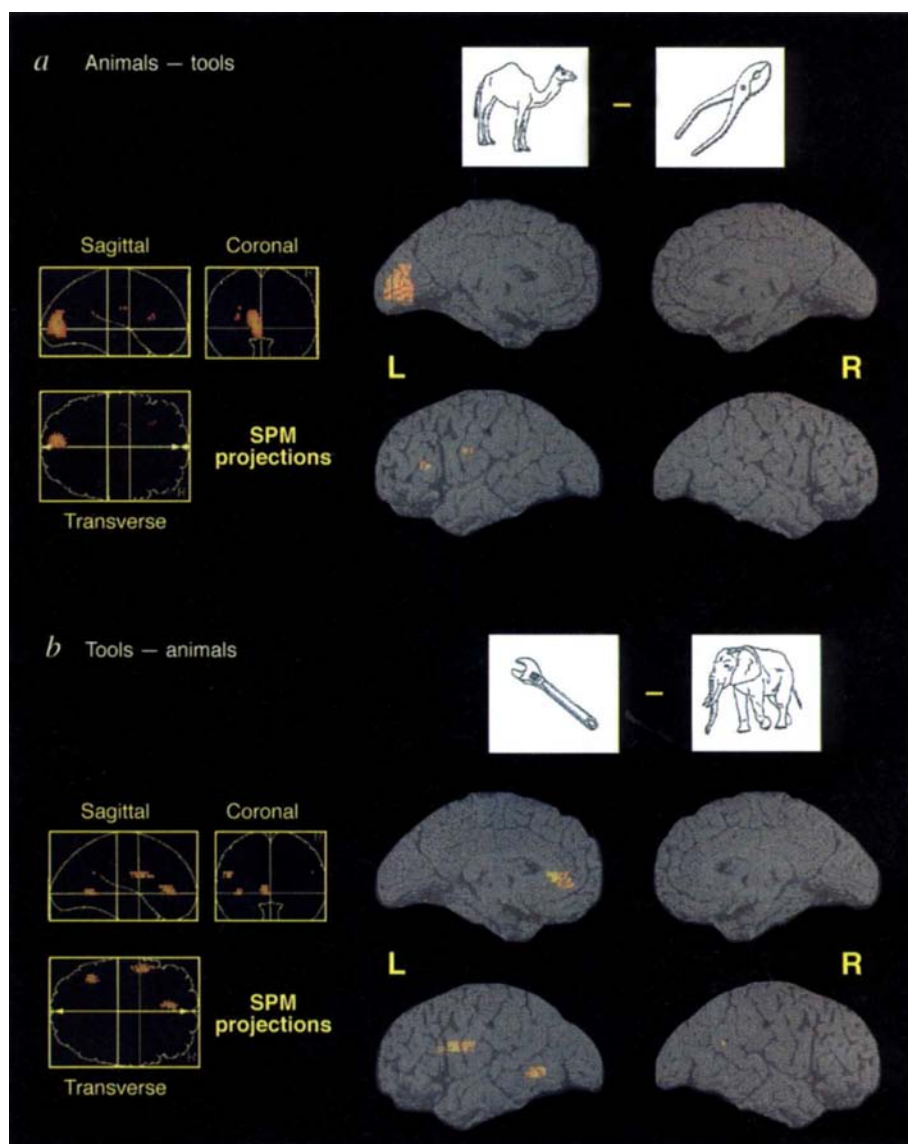
The region of the left middle temporal gyrus active when identifying tools, but not animals, was nearly identical to the area activated in previous studies in which subjects generated action words associated with objects^{8–10}. Thus, this region, which lies just anterior to the area active when perceiving movement^{17,18}, may be the site for stored knowledge about patterns of visual motion associated with using objects. The region of the left premotor cortex activated when subjects named tools, but not animals, was nearly identical to the area activated when subjects

imagined grasping objects with their right hand⁷. This premotor area, therefore, may be the site for stored knowledge about how objects are used. Both the left middle temporal gyrus and the left premotor cortex were also selectively activated by naming tool silhouettes relative to naming animal silhouettes (Fig. 3 legend). Thus, identifying tools may be mediated, in part, by areas that mediate knowledge of object motion, and of object use. Moreover, these representations appear to be stored close to the tissue that is active when perceiving motion and when using objects¹⁰.

Although patients with category-specific deficits typically have large lesions, difficulty naming living things has been associated with posterior and ventral lesions, whereas selective difficulty naming man-made objects has been associated with lesions that are anterior and dorsal¹⁹. Our findings are in agreement with this broad anatomical distinction but still provide a level of detail not previously available from studies on brain-damaged patients.

We have the capacity to identify a tremendous number of objects quickly and accurately. This ability is dependent, in part, on the automatic activation of previously acquired knowledge about the physical and functional attributes that define these

FIG. 3 Regions of increased rCBF when subjects silently named drawings of animals compared to silently naming tools (a), and when they silently named tools compared to silently naming animals (b). The strength of these activations were not as great as when each naming condition was compared to the nonsense object baseline condition (Fig. 2). To illustrate the spatial extent of these activations the threshold was set at $Z > 2.32$ ($P < 0.01$, 1-tailed), whereas the threshold for the activations depicted in Figs 1 and 2 was set at $Z > 3.09$, ($P < 0.001$, 1-tailed). Regions activated by animal naming compared to tool naming (a) were in the calcarine sulcus ($-4, -80, +8$; $Z = 3.12$). There were also two small activations in the left frontal lobe ($-26, -6, +24$; $Z = 2.50$, and $-26, +28, +16$; $Z = 2.41$). Regions activated by tool naming compared to animal naming (b) were in the left middle temporal gyrus ($-36, -50, +4$; $Z = 2.90$), left anterior cingulate (BA 32) ($-6, -38, +2$; $Z = 2.78$), right supramarginal gyrus ($+48, -50, +24$; $Z = 2.50$), and left lateral inferior frontal cortex, extending from BA 44 ($-52, +10, +20$; $Z = 2.85$) to premotor cortex (BA 6) ($-48, 0, +20$; $Z = 2.74$). Silhouette study: The voice response time was 703 ± 31.0 ms and error rate was $17.2 \pm 3.01\%$ for naming silhouettes of animals. The voice response time was 679 ± 24.4 ms and error rate was $15.6 \pm 4.05\%$ for naming tool silhouettes ($P > 0.10$). A region-of-interest analysis was applied to the rCBF data based on all pixels that exceeded a threshold of $Z = 2.32$ ($P < 0.01$) from the first study. Analysis of these regions indicated that, relative to naming silhouettes of tools, naming animal silhouettes activated the left medial occipital region ($F(1, 15) = 3.96$, $P < 0.05$) with maximal peak activity at $-18, -90, +12$ ($Z = 2.86$) and $-2, -84, -4$ ($Z = 2.63$), whereas, relative to naming animal silhouettes, tool silhouette naming activated the left middle temporal region ($F(1, 15) = 4.29$, $P < 0.05$) with maximal peak activity at $-40, -62, +4$ ($Z = 3.51$), and the left premotor region ($F(1, 15) = 8.43$, $P < 0.005$) with maximal peak activity at $-42, 0, +20$; $Z = 2.59$).



objects^{20,21}. Our results suggest that semantic representations of objects are stored as a distributed neural network that includes the ventral region of the temporal lobe. The location of the other areas recruited as part of this network depends on the intrinsic properties of the object to be identified. □

Received 11 July; accepted 1 December 1995.

- Warrington, E. K. & Shallice, T. *Brain* **107**, 829–854 (1984).
- Silveri, M. C. & Gainotti, G. *Cog. Neuropsychol.* **5**, 677–709 (1988).
- Sheridan, J. & Humphreys, G. W. *Cog. Neuropsychol.* **10**, 143–184 (1993).
- Warrington, E. K. & McCarthy, R. *Brain* **106**, 859–878 (1983).
- Warrington, E. K. & McCarthy, R. *Brain* **110**, 1273–1296 (1987).
- Sacchetti, C. & Humphreys, G. W. *Cog. Neuropsychol.* **9**, 73–86 (1992).
- Decety, J. et al. *Nature* **371**, 600–602 (1994).
- Wise, R. et al. *Brain* **114**, 1803–1817 (1991).
- Raichle, M. E. et al. *Cereb. Cortex* **4**, 8–26 (1994).
- Martin, A. et al. *Science* **270**, 102–105 (1995).
- Snodgrass, J. G. & Vanderwart, M. J. *exp. Psychol.: Hum. Learn. Mem.* **6**, 174–215 (1980).
- Kroll, J. F. & Potter, M. C. *J. verb. Learn. verb. Behav.* **23**, 39–66 (1984).
- Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, MA, 1982).
- Damasio, A. R. *Cognition* **33**, 25–62 (1989).
- Farah, M. J. *Trends Neurosci.* **12**, 395–399 (1989).
- Kosslyn, S. M. et al. *J. Cogn. Neurosci.* **5**, 263–287 (1993).
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L. & Petersen, S. *Science* **248**, 1556–1559 (1990).
- Zeki, S. et al. *J. Neurosci.* **11**, 641–649 (1991).
- Saffran, E. M. & Schwartz, M. F. in *Attention and Performance XV* (eds Umiltà, C. & Moscovitch, M.) 507–536 (MIT Press, MA, 1994).
- Glaser, W. R. *Cognition* **42**, 61–105 (1992).
- Humphreys, G. W., Riddoch, M. J. & Quinlan, P. T. *Cog. Neuropsychol.* **5**, 67–103 (1988).
- Tuving, E., Markovitsch, Kapur, S., Habib, R. & Houle, S. *NeuroReport* **5**, 2525–2528 (1994).
- Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme, New York, 1988).
- Friston, K. J. et al. *J. cereb. Blood Flow Metab.* **10**, 458–466 (1990).
- Friston, K. J. et al. *J. cereb. Blood Flow Metab.* **9**, 690–699 (1991).
- Friston, K. J., Frith, C. D., Liddle, P. F. & Frackowiack, R. S. J. *J. comput. assist. Tomogr.* **15**, 634–639 (1991).

Creation of a biologically active interleukin-5 monomer

Richard R. Dickason & David P. Huston

Baylor College of Medicine, Departments of Medicine, and Microbiology and Immunology, One Baylor Plaza, FBRN B567, Houston, Texas 77030, USA

INTERLEUKIN-5 (IL-5) specifically induces the differentiation of eosinophils, which are important in host defence and the pathogenesis of allergies and asthma^{1,2}. Structurally, IL-5 is a unique member of the short-chain helical-bundle subfamily of cytokines whose canonical motif contains four helices (A–D) arranged in an up–up–down–down topology^{3,4}. In contrast to other subfamily members, which fold unimolecularly into a single helical bundle^{5–8}, IL-5 forms a pair of helical bundles by the interdigitation of two identical monomers that contribute a D helix to the other's A–C helices³. We predicted that the lack of bioactivity by an IL-5 monomer⁹ was due to a short loop between helices C and D which physically prevents unimolecular folding of helix D into a functionally obligate structural motif. Here we report that, by lengthening this loop, we have engineered an insertional mutant of IL-5 that was expressed as a monomer with biological activity similar to that of native IL-5. These studies demonstrate that all of the structural features necessary for IL-5 to function are contained within a single helical bundle.

The consensus motif of helical-bundle cytokines requires the connection of helices A and B and helices C and D by long overhand loops (loops 1 and 3), and the connection of helices B and C by a short turn (loop 2)⁴. To determine the structural element(s) of interleukin-5 (IL-5) that influence its unique interdigitating homodimeric configuration, we compared the primary sequences of the secondary structural elements within IL-5 with the four other short-chain helical-bundle cytokines, these being IL-2, IL-4, granulocyte–macrophage–colony-stimulating factor

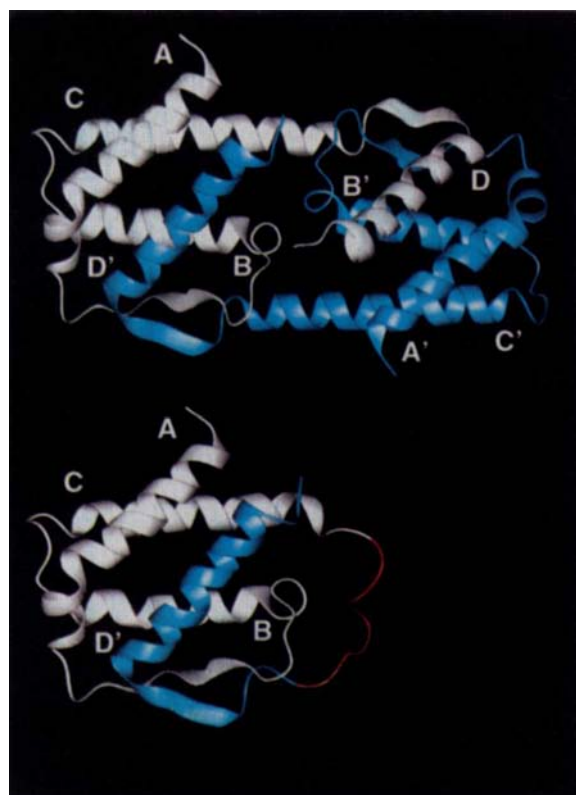


FIG. 1 Modelling of mono5. The top ribbon diagram represents the IL-5 interdigitating homodimer in which one monomer, with helices labelled A–D, is white and the other monomer, with helices labelled A'–D', is blue. Letters are located at the N-terminal end of each helix. The bottom ribbon diagram represents the modelled structure of mono5. This composite structure consists of residues 1–84 of one IL-5 monomer (white), residues 85–115 of the other IL-5 monomer (blue), and an insert (red) from residues 87–94 of the analogous loop 3 region of GM–CSF (His–Cys–Pro–Pro–Thr–Pro–Glu–Thr). This GM–CSF sequence contained three prolines, which favoured the extended secondary structure required to span the existing gap and did not perturb the helical-bundle conformation. To avoid aberrant formation of disulphide bonds, the cysteine at position 2 of the GM–CSF residues was changed to serine. The sequence for mono5 was constructed by replacing the *EaeI* and *BspMI* fragment of a previously reported human IL-5 construct¹¹ with a synthetic double-stranded insert. The insert encoded the native IL-5 sequence between the *EaeI* and *BspMI* sites, in addition to the eight codons of the modified GM–CSF loop 3 sequence. These eight codons were positioned between the native IL-5 codons for Lys 84 and Lys 85. Structural analysis of the short-chain helical-bundle cytokines used coordinates deposited in the Brookhaven Protein Databank (IL-2, 3ink⁵; IL-4, 1rcb⁶; GM-CSF, 1gmf⁷; and M-CSF, 1hmc⁸). The IL-5 coordinates were provided by S. Jordan³. The structures were displayed, manipulated and minimized on a Silicon Graphics workstation using the QUANTA software package (Keck Center for Computational Biology).

(GM–CSF), and M-CSF^{3,5–8}. IL-5 was distinguished by a shorter loop 3, containing only eight amino acids, which differed significantly from the 15–20 residues for the other short-chain helical-bundle cytokines. We therefore considered that the length of loop 3 might determine whether a helical-bundle cytokine will be actively expressed as a monomer or as an interdigitating homodimer. Molecular modelling suggested that lengthening loop 3 of IL-5 by eight amino acids would relieve the intramolecular folding constraint and enable unimolecular formation of a complete, and potentially functional, helical bundle.

A computer-generated image of the tertiary structure of IL-5 was modified by deletion of residues C-terminal to Lys 84 of one monomer, and residues N-terminal to Lys 85 of the other monomer. This targeted a site for insertional mutagenesis that preserved the stabilizing interhelical interactions, the β -strand

interactions, and the anchoring disulphide bond^{3,9} (Fig. 1). This modelled structure contained a gap, which was bridged by eight amino acids from the analogous loop 3 region of GM-CSF⁷. The nucleotide sequence coding for this insertional mutant of IL-5 (designated mono5) was engineered for expression in baculovirus-infected Sf-9 cells^{10,11}.

Affinity-purified mono5 was compared to similarly generated, native IL-5 (ref. 11) by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2, top) and western blot analysis (Fig. 2,

bottom). The migration of native IL-5 indicated that it has a relative molecular mass of 30,000 (M_r 30K) under non-reducing conditions (lane 1) and 15K under reducing conditions (lane 3). The migration of mono5 indicated on M_r of 16K under both non-reducing and reducing conditions, demonstrating that it is indeed a monomer. IL-5-specific monoclonal antibodies used on the western blot detected both IL-5 and mono5 under non-reducing conditions (lanes 1 and 2), but detected only mono5 under reducing conditions (lanes 3 and 4). Because these monoclonal

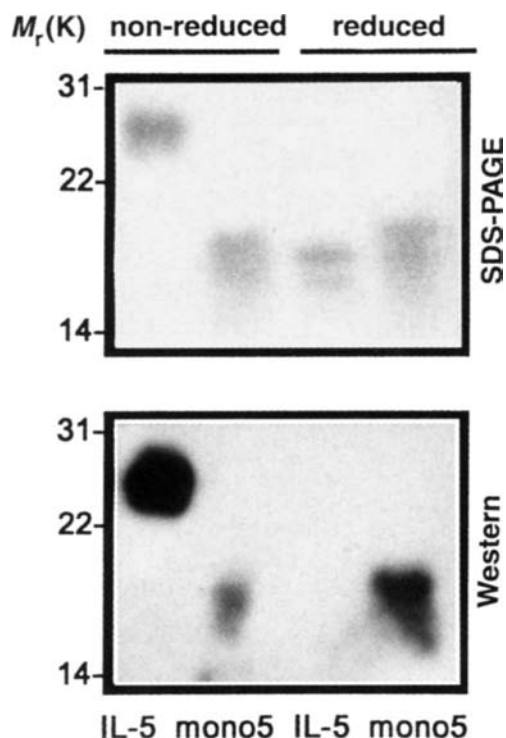


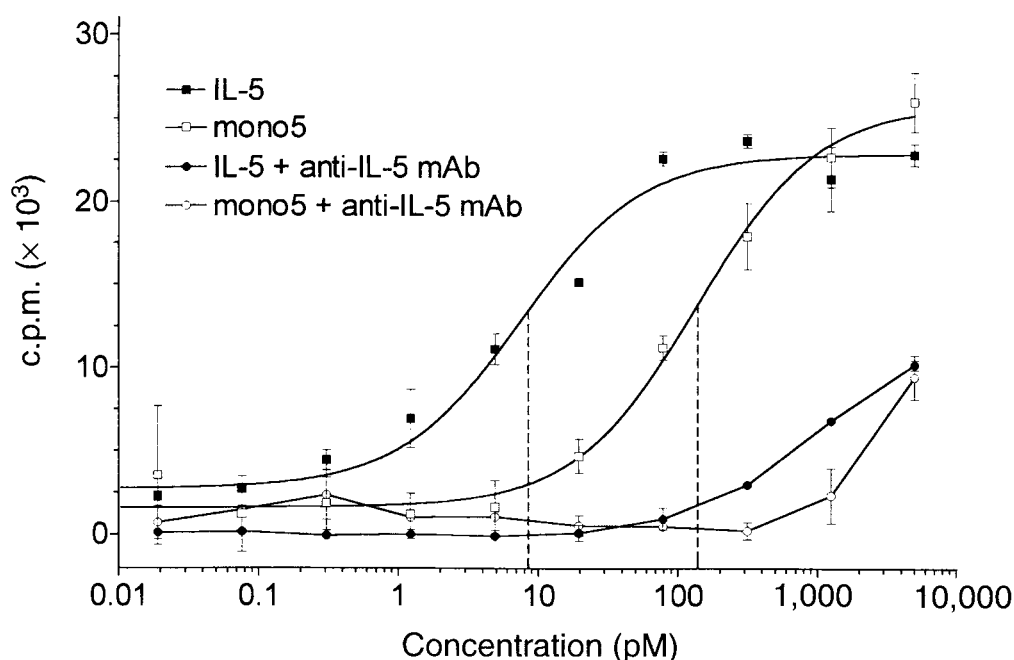
FIG. 2 Characterization of IL-5 and mono5 by SDS-PAGE and western blot analysis. Top, 200 ng recombinant affinity-purified IL-5 (lanes 1 and 3) or mono5 (lanes 2 and 4) were electrophoresed through an 8% polyacrylamide gel in the absence (lanes 1 and 2) or presence (lanes 3 and 4) of 2-mercaptoethanol following previous methods¹⁹ and silver stained. Bottom, recombinant IL-5 and mono5 were electrophoresed under the above conditions, blotted to a polyvinylidene fluoride membrane, and probed with the anti-human IL-5 monoclonal antibodies (mAbs) h5.2.4 and h5.7.3 ($2 \mu\text{g ml}^{-1}$) as previously described¹¹. Numbers to the left of each panel denote relative molecular mass (M_r).

METHODS. The mono5 sequence was inserted into the baculovirus transfer vector pV1393 and recombinant virus generated as previously described¹¹. Sf-9 cells were infected and recombinant mono5 protein was expressed following standard protocols¹⁰. Baculovirus-expressed recombinant mono5 was purified over an anti-human IL-5 mAb affinity column and quantified using the bicinchoninic acid protein assay as previously described¹¹. The predicted M_r for mono5 was determined by halving 30K, the M_r of the native IL-5 dimer, and adding 1K, the M_r of the insert of eight amino acids.

FIG. 3 Bioactivity of IL-5 and mono5. IL-5 (filled symbols) and mono5 (open symbols) were titrated in the TF-1 proliferation assay in the absence (squares) or presence (circles) of h5.2.4, a human IL-5-specific neutralizing mAb¹¹. The vertical broken lines intersect the x-axis at the effector concentration for half maximum proliferative response (EC_{50}) for IL-5 ($9 \pm 1 \text{ pM}$) and mono5 ($139 \pm 12 \text{ pM}$). In addition, mono5 was assayed for its capacity to induce proliferation of a subline of TF-1 cells that is responsive to GM-CSF but not IL-5. Lack of mono5 bioactivity with this subline (data not shown) demonstrates that residues in the GM-CSF insert did not impart any GM-CSF-specific bioactivity to mono5.

METHODS. A stock of the IL-5-responsive human erythroleukemic TF-1 cells was provided by T. Kitamura¹². The IL-5-non-responsive TF-1 subline was obtained from the American Type Culture Collection (CRL 2003).

Fourfold serial dilutions of baculovirus-expressed affinity-purified IL-5 or mono5, starting at 5 nM, were incubated with 20,000 TF-1 cells in duplicate wells. Cells were incubated for 48 h and were labelled for the



last 12 h with ^3H -thymidine. ^3H -thymidine incorporation from collected cells was assessed by scintillation counting. The EC_{50} was calculated using the PRISM software package.

antibodies bind conformational epitopes on native IL-5 (ref. 11), mono5 must assume a conformation which retains these epitopes even in the absence of an intramolecular disulphide bond.

The bioactivity of mono5 was demonstrated by its capacity to induce proliferation (Fig. 3) of the IL-5-responsive TF-1 cell line¹², and its maximum bioactivity was similar to that of native IL-5. IL-5-specific monoclonal antibodies also neutralized the activity of both IL-5 and mono5, indicating that the proliferative responses were specific to IL-5. The concentration of IL-5 and mono5 which induced half-maximum and maximum proliferation of TF-1 cells differed by approximately one order of magnitude (Fig. 3). These differences probably reflect a lower affinity of mono5 for its receptor complex, which could be accounted for by a slight perturbation of functional residues induced by the GM-CSF insert. Alternatively, the GM-CSF insert may present some steric hindrance to the engagement of mono5 with the IL-5 receptor (IL-5R). In addition, mono5 may lack accessory residues present on the native IL-5 homodimer which are not necessary for function but enhance affinity for IL-5R.

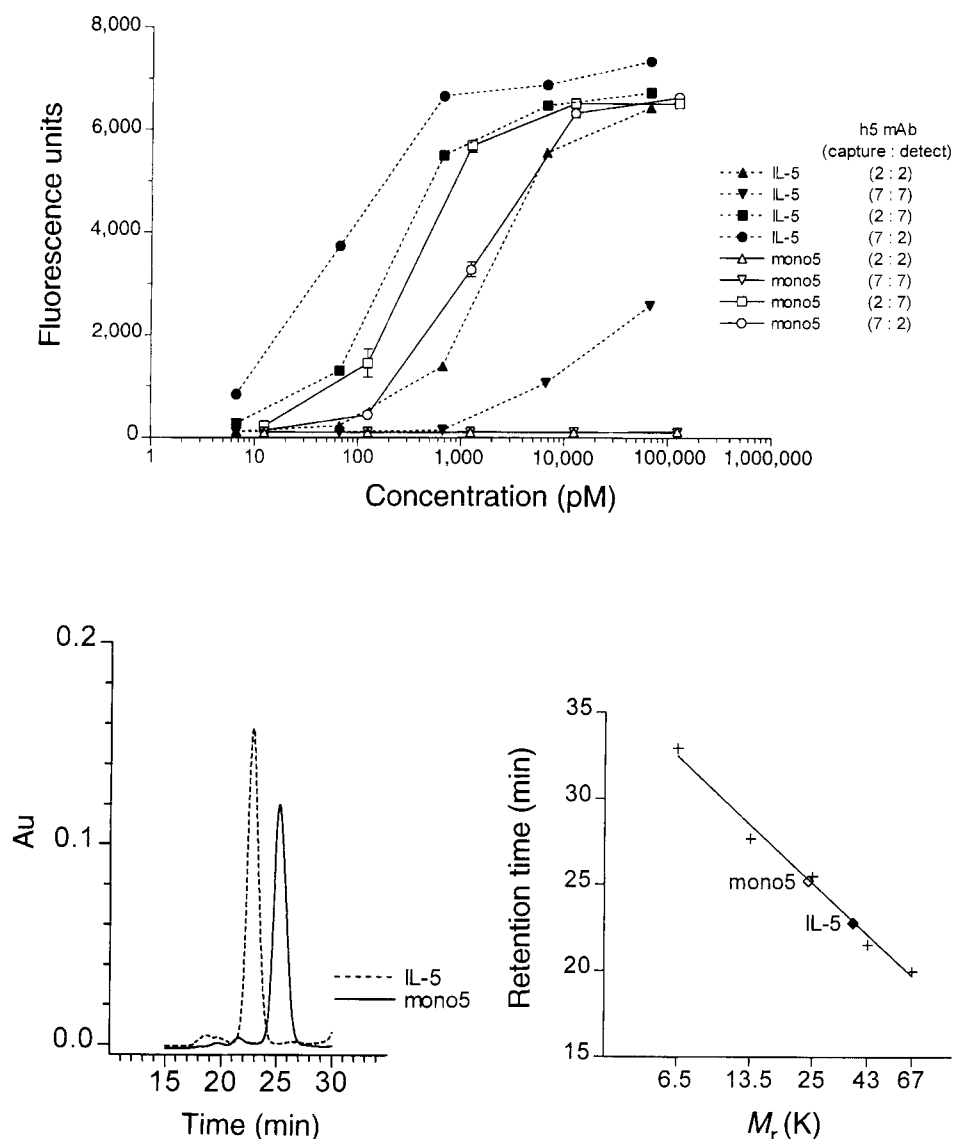
To ensure that there was no non-covalent association of mono5 molecules, two non-denaturing analyses were performed. By sandwich enzyme-linked immunosorbent assay (ELISA), the IL-5 homodimer displayed paired epitopes that were detected by using the same monoclonal antibody to both capture and detect.¹¹ By contrast, mono5 was only detected in sandwich ELISA by using two non-competitive monoclonal antibodies (Fig. 4, top), indi-

cating that only single epitopes were present. Size-exclusion high-performance liquid chromatography (HPLC) also indicated that mono5 exists only as a monomeric species in solution (Fig. 4, bottom). Computer modelling further demonstrated that the insert of eight amino acids, by displacing a volume not present on the native IL-5 dimer, sterically hinders two mono5 molecules from mimicking the proximity and orientation of the paired helical bundles of native IL-5 for IL-5R engagement. Although the possibility that signal transduction could be dependent on the binding of two monomers to one IL-5R has not been formally excluded, these studies do indicate that there is no intermolecular association of mono5. Furthermore, the capacity of mono5 to have maximum bioactivity equivalent to that of IL-5 clearly demonstrates that all of the structural features necessary for IL-5 to function are contained within a single helical bundle.

The effects of the length of loop 3 on the configuration of IL-5 also provide a theoretical basis for the creation of other new monomeric and homodimeric helical bundle proteins. Our structural hypothesis leads us to predict that deletion of a portion of loop 3 in a monomeric helical-bundle cytokine (for example, GM-CSF) might necessitate the formation of an interdigitating homodimer for completion of a functional helical bundle.

The IL-5R is composed of a cytokine-specific α -chain (IL-5R- α) and a signal-transducing β -chain, which is shared with the cytokine-specific α -chains for GM-CSF and IL-3 (refs 13, 14). The IL-5R- α binding domain of IL-5 has been localized to the junction between

FIG. 4 Characterization of IL-5 and mono5 under non-denaturing conditions. Top, sandwich ELISAs were performed as previously described¹¹ with both IL-5 (filled symbols) and mono5 (open symbols), using the mAbs h5.2.4 and h5.7.3 in the following four combinations of capture mAb: detection mAb; h5.2.4:h5.2.4 (upright triangles); h5.2.4:h5.7.3 (squares); h5.7.3:h5.7.3 (inverted triangles); and h5.7.3:h5.2.4 (circles). Both IL-5 and mono5 were detected using non-competing mAb pairs (squares and circles). In contrast to IL-5, which was also detected when the same mAb was used to capture and detect (filled upright and inverted triangles), mono5 could not be detected when the same mAb was used to capture and detect (open upright and inverted triangles, which are superimposed along the baseline at all concentrations). Bottom, 2 μ g IL-5 and 1 μ g mono5 were analysed by size-exclusion HPLC (with PBS, pH 7.4, at room temperature). The Superdex-75 column was calibrated with known M_r standards with retention times as follows: 6.5K, 33.0 min; 13.5K, 27.7 min; 25K, 25.5 min; 43K, 21.5 min; and 67K, 20.0 min (bottom right, crosses). IL-5 migrated as a single species (bottom left, dotted line) with a retention time of 22.8 min, which equated to the approximate M_r for an IL-5 homodimer (bottom right, filled diamond). Mono5 also migrated as a single species (bottom left, solid line) but with a distinct retention time of 25.2 min, which equated to the approximate M_r for a non-oligomerized mono5 (bottom right, open diamond). Mono5 was only detected by sandwich ELISA in fractions from the 25.2-min peak, and these same fractions were the only samples to possess TF-1 bioactivity ($EC_{50} = 141 \pm 11$ pM). These studies demonstrate that intermolecular association of mono5 does not occur.



loop 3 and helix D and the N-terminal end of helix D; binding of the β -chain has been localized to the middle of helix A¹⁵⁻¹⁷. However, the role of the IL-5 homodimeric topology for bioactivity remained unknown, because IL-5 binds IL-5R- α in a 1:1 stoichiometry¹⁵⁻¹⁸. Our studies eliminate the possibilities that a receptor binding domain spans both helical bundles of IL-5 and that paired helical bundles are essential for IL-5 to function. Furthermore, limiting the functional structure of IL-5 to a single helical bundle provides a framework for the rational design of IL-5 therapeutic analogues. □

Received 31 August; accepted 14 December 1995.

1. Sanderson, C. J. *Adv. Pharmac.* **23**, 163–177 (1992).
2. Korker, M. & Takatsu, K. *Int. Archs Allergy appl. Immun.* **104**, 1–9 (1994).
3. Milburn, M. V. *et al. Nature* **363**, 172–176 (1993).

4. Rozwarski, D. A. *et al. Structure* **2**, 159–173 (1994).
5. Bazan, J. F. & McKay, D. B. *Science* **257**, 410–418 (1992).
6. Smith, L. J. *J. molec. Biol.* **224**, 899–904 (1992).
7. Diederichs, K., Boone, T. & Karplus, A. *Science* **254**, 1779–1782 (1991).
8. Pandit, J. *et al. Science* **258**, 1358–1362 (1992).
9. McKenzie, A. N., Ely, B. & Sanderson, C. J. *Molec. Immun.* **28**, 155–158 (1991).
10. Summers, M. D. & Smith, G. E. *A Manual of Methods for Baculovirus Vector and Insect Cell Culture Procedures* (Texas Agriculture Experiment Station, Texas, 1987).
11. Dickason, R. R., Huston, M. M. & Huston, D. P. *Cytokine* **6**, 647–656 (1994).
12. Kitamura, T. *et al. J. Cell Physiol.* **140**, 323–334 (1989).
13. Tavernier, J. *et al. Cell* **65**, 1175–1184 (1991).
14. Kitamura, T., Sato, N., Arai, K. & Miyajima, A. *Cell* **65**, 1165–1174 (1991).
15. Tavernier, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 5194–5198 (1995).
16. Graber, P. *et al. J. biol. Chem.* **270**, 15762–15769 (1995).
17. Cornelis, S. *et al. EMBO J.* **14**, 3395–3402 (1995).
18. Devos, R. *et al. J. biol. Chem.* **268**, 6581–6587 (1993).
19. Laemmli, U. K. *Nature* **227**, 680–685 (1970).

ACKNOWLEDGEMENTS. We thank R. Cook, M. Estes, J. Guevara and J. Morrisett for advice and assistance; S. Crawford, J. Moore and G. Tavana for technical assistance; and M. Entman, H. Gilbert, G. Harriman, M. Huston, J. Morrisett and R. Rich for critical reviews of this manuscript.

Identification of the *gal4* suppressor Sug1 as a subunit of the yeast 26S proteasome

David M. Rubin*, Olivier Coux*, Inge Wefes*, Christoph Hengartner†, Richard A. Young†, Alfred L. Goldberg* & Daniel Finley*‡

* Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA
† Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

THE *SUG1* gene of *Saccharomyces cerevisiae* encodes a putative ATPase. Mutations in *SUG1* were isolated¹ as suppressors of a mutation in the transcriptional activation domain of *GAL4*. *Sug1* was recently proposed to be a subunit of the RNA polymerase II holoenzyme and to mediate the association of transcriptional activators with holoenzyme². We show here that *Sug1* is not a subunit of the holoenzyme, at least in its purified form, but of the 26S proteasome^{3,4}, a large complex of relative molecular-mass 2,000K that catalyses the ATP-dependent degradation of ubiquitin–protein conjugates. *Sug1* co-purifies with the proteasome in both conventional and nickel-chelate affinity chromatography. Our observations account for the reduced ubiquitin-dependent proteolysis in *sug1* mutants⁵ and suggest that the effects of *sug1* mutations on transcription are indirect results of defective proteolysis.

The composition of the 26S proteasome from *S. cerevisiae* has been the subject of controversy^{2,5}, owing largely to the difficulty of obtaining pure preparations of the 26S proteasome from this organism. To purify this complex, we first fractionated yeast extracts on DEAE–Affigel blue. The peak of hydrolytic activity against succinyl-Leu-Leu-Val-Tyr-7-amino-4-methyl coumarin (Suc-LLVY-AMC) was coincident with the peak of activity against ubiquitin–lysozyme conjugates (results not shown). The most active fractions were further purified by chromatography on Mono-Q (Fig. 1a–c) and Superose-6 columns (Fig. 1d–f). The final preparations were nearly homogeneous by electron microscopy (Z. Cejka and W. Baumeister, personal communication), and the subunit composition of the yeast 26S proteasome on SDS–PAGE (Fig. 1e) closely resembled that of mammals^{6,7}.

Throughout the purification, *Sug1*, identified by immunoblotting, cofractionated with the 26S proteasome (Fig. 1c,f). The chromatographic behaviour of *Sug1* closely paralleled that of *Cim5*, a related ATPase, whose human homologue has, like *Sug1*, been implicated in transcription⁸ and protein degradation⁹ (Fig. 1c,f). A particular concern was to rule out possible contamination of the 26S proteasome preparation by RNA polymerase II holoenzyme¹⁰, which has been proposed¹¹ to contain *Sug1*. Therefore the presence of holoenzyme was monitored throughout the purification by immunoblotting, using antibodies¹² against holoenzyme subunits *Srb4* and *Srb5*. The 26S proteasome was effectively separated from holoenzyme during the DEAE–Affigel blue step (not shown). Any residual holoenzyme in the eluate was eliminated in the Mono-Q fractionation (Fig. 1c).

It has been suggested² that the 26S proteasome lacks *Sug1* but contains a closely related protein that crossreacts with anti-*Sug1* antibodies. To test this possibility, we tagged *Sug1* with a His₆ epitope¹³. When the 26S proteasome was purified from cells expressing His₆-*Sug1*, the protein detected by the anti-*Sug1* antibody had reduced electrophoretic mobility, demonstrating that it is the *SUG1* gene product (Fig. 1g).

To confirm the presence of *Sug1* in the 26S proteasome, we developed an independent method for purification of the 26S proteasome. A His₆ epitope was fused to Pre1, a subunit of the 20S core particle¹⁴ of the 26S proteasome⁴. Extracts were prepared from cells expressing either His₆-Pre1 or wild-type Pre1, and fractionated by nickel-chelate affinity chromatography¹⁵. Material that bound specifically to the column was eluted with 100 mM imidazole. The eluate from the tagged preparation contained high levels of peptidase and ubiquitin-conjugate-degrading activity, unlike that from wild-type cells (Fig. 2a). Both *Sug1* and *Cim5* proteins were enriched in the column eluate prepared from His₆-Pre1 cells but not in the eluate from wild-type cells (Fig. 2b).

These results demonstrate that *Sug1* is an integral component of the 26S proteasome. But several studies^{2,11,15} have proposed that *Sug1* is a subunit of RNA polymerase II holoenzyme and mediates the physical interaction of holoenzyme with transcriptional activators such as *Gal4*. We tested this model by immunoblot analysis of a purified, previously characterized preparation¹⁶ of holoenzyme. Recombinant *Sug1* was used as a standard to estimate the number of copies of *Sug1* per holoenzyme. The 1:1 stoichiometry expected for a subunit of holoenzyme was not found. Rather, the number of copies of *Sug1* per holoenzyme appeared to be less than 0.05:1, comparable to our limit of detection (Fig. 3a). In contrast, when the 26S proteasome was immunoblotted under identical conditions, *Sug1* appeared to be present at about one copy per 26S particle (Fig. 3b). Although we cannot exclude the possibility that some *Sug1* reversibly associates with holoenzyme or associates with a minor subfraction of holoenzyme, the set of proteins that constitute holoenzyme in its purified and active form does not appear to include *Sug1*.

‡ To whom correspondence should be addressed.

Our data, together with previous studies on both yeast and mammals^{5,9,17,18}, call into question the notion that Sug1 and the related ATPases Mss1, Tbp1 and Tbp7 participate directly in the control of transcription^{1,2,11,15,19–21}. The argument that Sug1 is a transcription factor was originally based on the ability of *sug1* mutations to suppress the transcriptional defect of a *gal4* activation-domain deletion¹ (*gal4D*), but this study did not examine possible effects of Sug1 on turnover of the mutant Gal4D protein. Interestingly, suppressors of mutations in both the Gal4 and Cdc68 transcription factors appear to arise much more frequently in *SUG1* than in genes encoding other ATPases of the 26S

proteasome^{1,22}. The basis for selectivity towards Sug1 as opposed to other proteasomal ATPases remains unclear. Phenotypic differences among proteasomal ATPases might reflect the specificity of physical interactions between protein substrates and the ATPases. Thus, the multiplicity of such ATPases may endow the 26S proteasome with a capacity to interact with and degrade a diverse set of ubiquitinated proteins. The precise role of ATP hydrolysis in the mechanism of protein breakdown has not been identified, but may be linked to substrate unfolding and translocation into the proteolytic lumen of the 26S proteasome^{4,23,24}. □

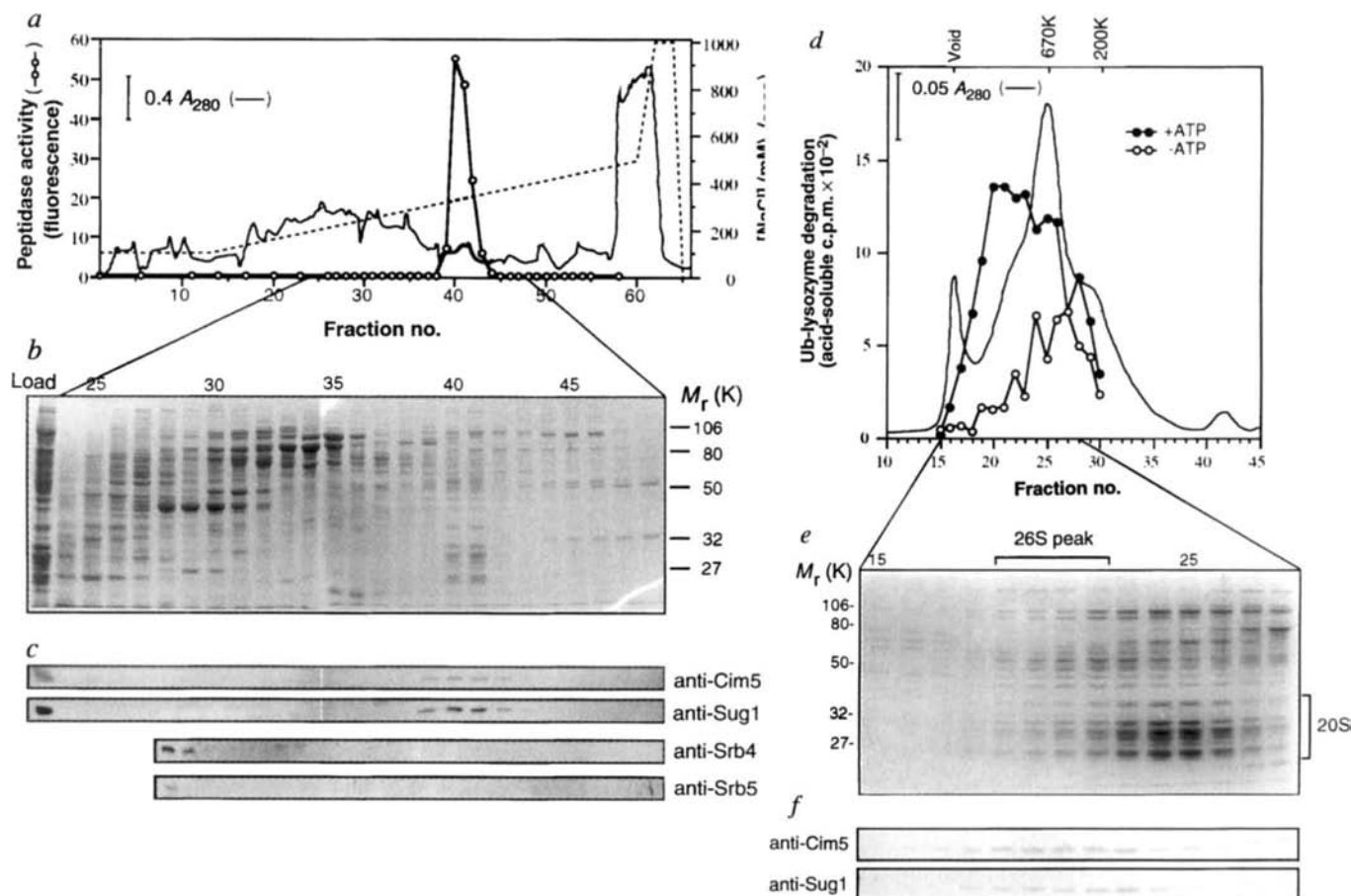


FIG. 1 Copurification of the Sug1 gene product with the 26S proteasome. *a–c*, Analysis of Mono-Q fractions. *a*, Circles represent peptidase activity against Suc-LLVY-AMC. *b*, 12% SDS-PAGE analysis of column fractions, stained with Coomassie. *c*, Immunoblots of column fractions probed with antibodies to the proteins indicated. *d–f*, Superose-6 fractionation of the Mono-Q eluate (fractions 40 and 41). *d*, Circles, degradation of ubiquitin-lysosome conjugates in the presence and absence of ATP. *e*, 12% SDS-PAGE analysis of column fractions, stained with Coomassie. Bracket on the right indicates the protein bands from the 20S subcomplex of the 26S proteasome¹⁴. Bracket at the top indicates the 26S proteasome-containing fractions. These fractions contained the highest ATP-dependent ubiquitin-conjugate-degrading activity. In addition, peptidase activity in these fractions (not shown) was inhibited by SDS (unlike the activity in fractions 24–26, which contained free, SDS-activated²⁵ 20S particles). Fractions 24–26 contain 20S particles, as indicated by peptidase assay, ATP-independent degradation of ubiquitin conjugates and protein composition. *f*, Immunoblots of Superose-6 column fractions. *g*, Immunoblot analysis of wild-type and His₆-tagged Sug1 in purified 26S proteasomes. Samples from DY85 (WT) and DY177 (His₆-Sug1) cells were purified and characterized as for *a–f*. 26S proteasomes from DY177 were chromatographically indistinguishable from those of DY85, and the construct expressing His₆-tagged Sug1 fully complemented a *sug1* null mutation.

METHODS. Extracts were prepared from *S. cerevisiae* strain DY85 (used as

wildtype). DY85 is isogenic with SUB62 (ref. 26) except that it carries a chromosomal deletion of *CIM5* and a YCplac111-derived²⁷ plasmid containing *CIM5*. DY177 is isogenic with SUB62 except that it carries a chromosomal deletion of *SUG1* and a YCplac22-derived²⁷ plasmid expressing Sug1 tagged at its N terminus with His₆. Antibodies against Sug1 and Cim5 were provided by C. Mann⁵. The Srb immunoblots are from an independent, similarly executed purification; in other fractionations, Srb proteins were undetectable following the DEAE-Affigel Blue step. Immunodetection was performed using secondary antibodies coupled to alkaline phosphatase. Details of the purification are available as supplementary information on the Nature web site (<http://www.nature.com>).

FIG. 3 Immunoblot analysis of Sug1 levels in *a*, RNA polymerase II holoenzyme (Holo), and *b*, the 26S proteasome. Sug1 and Srb5 standards are purified recombinant proteins. Srb5 is a stoichiometric subunit of holoenzyme¹⁶. The asterisk marks the principal immunoreactive band in the holoenzyme sample, which does not comigrate with recombinant Sug1.

METHODS. The 26S proteasome sample corresponds to fractions 21 and 22 of Fig. 1d–f. Holoenzyme was purified and characterized as described¹⁶. A number of proteins in addition to Srb5 are present at one copy per holoenzyme complex in this specific preparation¹⁶. The molar concentration of 26S proteasome was estimated from the total protein concentration, assuming 100% purity and M_r 2×10^6 (ref. 6). Protein concentrations were measured using the Protein Assay reagent (Bio-Rad) with BSA as standard.

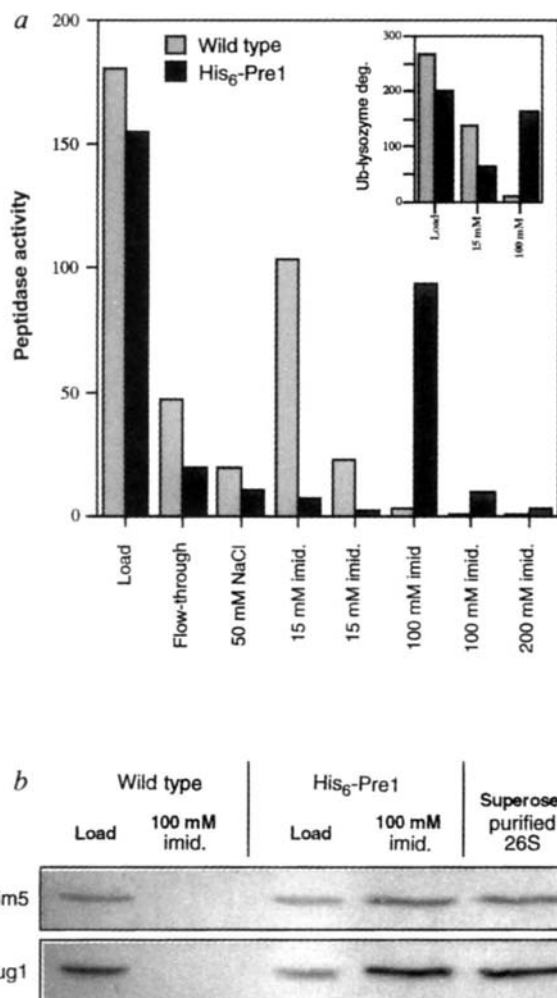
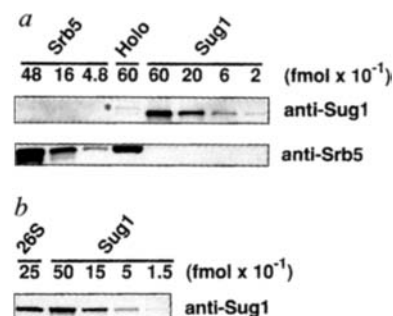


FIG. 2 Association of Sug1 with Pre1 shown by affinity chromatography. The Pre1 subunit of the 26S proteasome was tagged with a His₆ epitope. The His₆-Pre1 construct fully complemented a lethal *pre1* null mutation. Extracts from His₆-Pre1 and control strains were fractionated on Ni affinity columns. *a*, The epitope-tagged complex elutes at 100 mM imidazole (imid.), as indicated by either peptidase activity or (inset) ubiquitin-conjugate-degrading activity. Material was eluted from the column using a series of buffers (left to right). Activity values represent total activity in the fraction. *b*, Immunoblot of column fractions. Sug1 and Cim5 are detected in the 100 mM imidazole eluate of His₆-Pre1 but not wild-type extracts.

METHODS. Wild-type extract was from strain DY85; the His₆-Pre1 extract from SUB459, which is isogenic with YW071 (ref. 28) apart from the His₆ extension at the C terminus of Pre1. Extracts were processed as for Fig. 1 except that EDTA and DTT were omitted and Tris was replaced with sodium phosphate (50 mM) in all solutions. Eluates from the DEAE-Affigel blue column were loaded directly onto Ni-NTA (nitrilotriacetic acid) agarose columns. Nonspecifically bound proteins were eluted with a solution containing 50 mM NaCl, followed by a solution containing 15 mM imidazole and 100 mM NaCl. Equal proportions of the 100 mM imidazole eluate were immunoblotted for wild-type and 6HIS-Pre1 samples. Bands in the affinity eluate are more intense than in the load because a higher proportion of the eluate was immunoblotted.

Received 8 September; accepted 23 November 1995.

- Swaffield, J. C., Bromberg, J. F. & Johnston, S. A. *Nature* **357**, 698–700 (1992).
- Swaffield, J. C., Melcher, K. & Johnston, S. A. *Nature* **374**, 88–91 (1995).
- Dubiel, W., Ferrel, K. & Rechsteiner, M. *Molec. Biol. Rep.* **21**, 27–34 (1995).
- Rubin, D. M. & Finley, D. *Curr. Biol.* **5**, 854–858 (1995).
- Ghislain, M., Udvardy, A. & Mann, C. *Nature* **366**, 358–362 (1993).
- Yoshimura, T. et al. *J. struct. Biol.* **111**, 200–211 (1993).
- Hoffman, L., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **267**, 22362–22368 (1992).
- Shibuya, H. et al. *Nature* **357**, 700–702 (1992).
- Dubiel, W., Ferrel, K. & Rechsteiner, M. *FEBS Lett.* **323**, 276–278 (1993).
- Koleske, A. J. & Young, R. A. *Trends biochem. Sci.* **20**, 113–116 (1995).
- Kim, Y.-J., Björklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. *Cell* **77**, 599–608 (1994).
- Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. *Cell* **73**, 1361–1375 (1993).
- Crowe, J. et al. *Meth. molec. Biol.* **31**, 371–387 (1994).
- Heinemeyer, W., Kleinschmidt, J. A., Saidowsky, J., Escher, C. & Wolf, D. H. *EMBO J.* **10**, 555–562 (1991).
- Lee, J. W., Ryan, F., Swaffield, J. C., Johnston, S. A. & Moore, D. D. *Nature* **374**, 91–94 (1995).
- Koleske, A. J. & Young, R. A. *Nature* **368**, 466–469 (1995).
- Dubiel, W., Ferrel, K., Pratt, G. & Rechsteiner, M. *J. biol. Chem.* **267**, 22699–22702 (1992).
- Akiyama, K. et al. *FEBS Lett.* **363**, 151–156 (1995).
- Nelbock, P., Dillon, P. J., Perkins, A. & Rosen, C. A. *Science* **248**, 1650–1653 (1990).
- Ohana, B. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 138–142 (1993).
- Lee, J. W., Choi, H.-S., Gyuris, J., Brent, R. & Moore, D. D. *Molec. Endocr.* **9**, 243–254 (1995).
- Xu, Q., Singer, R. A. & Johnston, G. C. *Molec. cell. Biol.* **15**, 6025–6035 (1995).
- Lupas, A., Koster, A. J. & Baumeister, W. *Enz. Prot.* **47**, 252–273 (1993).
- Wenzel, T. & Baumeister, W. *Nature struct. Biol.* **2**, 199–203 (1995).
- Peters, J.-M., Franke, W. W. & Kleinschmidt, J. A. *J. biol. Chem.* **269**, 7709–7718 (1994).
- Finley, D., Özkaynak, E. & Varshavsky, A. *Cell* **48**, 1035–1046 (1987).
- Gietz, R. D. & Sugino, A. *Gene* **74**, 527–534 (1988).
- Seufert, W. & Jentsch, S. *EMBO J.* **11**, 3077–3080 (1992).

SUPPLEMENTARY INFORMATION is on www.nature.com. Paper copies are available from Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. D.M.R. and O.C. contributed equally to this work. We thank Z. Cjeka and W. Baumeister for permission to cite unpublished data, S. Jentsch and J. Callis for strains, D. Hwang for assistance with affinity purification of the 26S proteasome, and are especially grateful to C. Mann and M. Ghislain for advice and reagents. This work was supported by grants from the March of Dimes and the NIH to D.F., from the NIH and the Human Frontiers Science Program (HFSF) to A.L.G., by an NIH fellowship (D.M.R.), and a HSPF Fellowship (O.C.).

ERRATUM

Syk tyrosine kinase required for mouse viability and B-cell development

Alec M. Cheng, Bruce Rowley, William Pao, Adrian Hayday, Joseph B. Bolen & Tony Pawson

Nature **378**, 303–306 (1995)

In the legend to Fig. 4 of this Letter, the designations for rf2 and rf3 were confused: bold hatching represents rf2 and light hatching rf3.

CORRECTION

A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein

Jonathan C. Swaffield, Karsten Melcher
& Stephen Albert Johnston

Nature 374, 88–91 (1995)

In this Letter we reported that Sug1 protein was not a component of the 26S proteasome, based on our finding that Sug1p did not significantly co-fractionate or co-immunoprecipitate with subunits of the 20S proteasome. In these experiments we used a strain in which *SUG1* was deleted and an epitope-tagged *SUG1* was carried on a plasmid. We have found that under these conditions tagged Sug1 protein accumulates to abnormally high levels. The excess Sug1 protein does not associate with the 20S proteasome, but associates instead with DNA and TBP as reported. However, we have since found that when expressed at normal levels, Sug1 protein co-fractionates with the 20S proteasome (Fig. 1).

Equal amounts of whole-cell extract from wild-type S10-Sug1 (*S10-SUG1* reintegrated into the *SUG1* locus) and overexpressed S10-Sug1 (Δ SUG1 pVT100-U-*S10-SUG1*) strains were passed separately down a gel filtration column. Fractions were collected

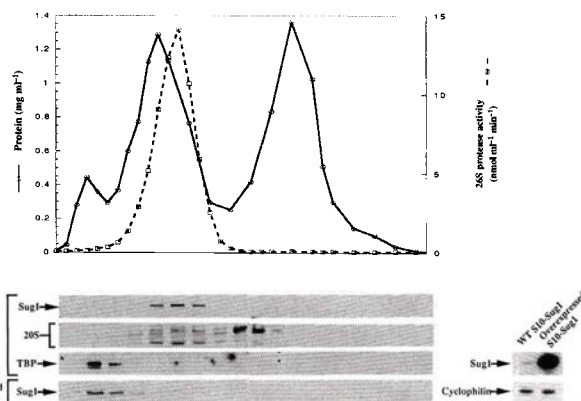


FIG. 1 Sug1 protein co-migrates with 20S proteasome subunits upon gel filtration in extracts from a wild-type (WT) strain but co-migrates with TBP from a strain overexpressing Sug1. Data include contribution from Steven J. Russell.

and assayed for Sug1 protein (using an antibody against the S10 epitope tag), 20S subunits and TBP by western blot. The protein and 26S activity plot is representative of the profiles obtained for both strains. S10-Sug1 levels were determined by western blot with anti-S10 antibody; cyclophilin served as a loading control. From this and our other data, we conclude that most of Sug1 protein is associated with the 26S proteasome when expressed at normal levels. □

KNOW YOUR COPY RIGHTS RESPECT OURS

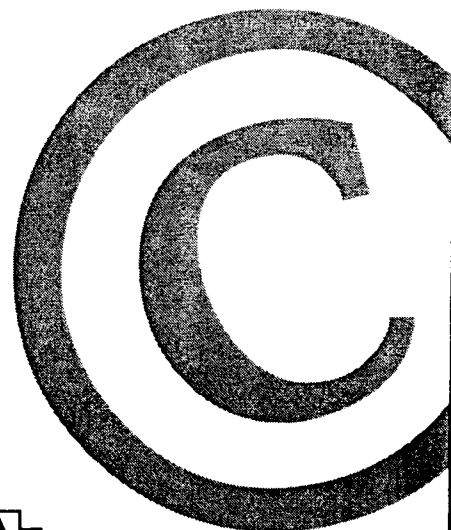
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

Chromatography column

Manufacturers show their true colours with new separation products, such as an HPLC solvent recycler, new HIC columns, a capillary GC injector, a microflow-HPLC system and a selection of chromatography data control systems.

SPARK HOLLAND has combined a high-resolution **syringe dispenser with a sampling system** that enables the Triathlon autosampler to inject 1 µl samples with a precision typically of better than 0.5 per cent relative standard deviation. The sampling system also allows the use of flexible PEEK tubing or a fused silica capillary as the sampling 'needle'. Aside from being inert and biocompatible, this needle enables injections from very small sample volumes. Additional Peltier-based sample cooling (an integrated option) and extensive reagent addition and mixing capabilities should make this autosampler a valuable tool.

Reader Enquiry No. 100

PhaseSeparations has recently released a new **vials catalogue** that covers autosampler consumable needs. It contains a range of assorted vials, caps, stoppers, seals and septa for all common auto-



The new PhaseSeparations catalogue.

samplers. Also offered are accessories, serum, storage and reaction bottles, and Hamilton autosampler syringes for GC and HPLC. The new catalogue also includes Chromacol and Wheaton vials, in addition to the company's own range manufactured from borosilicate type-1 glass.

Reader Enquiry No. 101

The new Model RC-200 **solvent recycler** from Phenomenex offers advanced features for automation and good manufacturing practice. This new model is said to work

with any isocratic system to recover and recycle automatically up to 80 per cent of the mobile phase typically shunted to waste. Some of the new features of the RC-200 recycler include automatic storage and recall of up to 10 methods, a universal power supply and compatibility with any detector. Detection of both positive and negative peaks is possible, with a selectable recycler signal display that can be matched to that of the detector. The recycler is now capable of validated output to document valve operation. Moreover, the new selfcleaning feature enables the unit automatically to wash/rinse the HPLC column and lines before shutting itself down.

Reader Enquiry No. 102

Media and columns

Pharmacia Biotech has expanded its range of Resource **prepacked columns** with Resource ETH (ether), ISO (isopropyl) and PHE (phenyl hydrophobic), three new columns for high-performance hydrophobic interaction chromatography (HIC). These columns are intended for the final stages of a purification process where trace contaminants and close variants must be removed. The media packed in these columns are based on Source 15-µm monosized beads. The manufacturer states that these HIC columns are characterized by high resolution that is well maintained at high loads and high flow rates. These columns should prove useful for many protein and peptide separation problems. The columns (1-ml bed volume) can be run with FPLC and HPLC systems, as well as with low to medium pressure systems and even a simple peristaltic pump.

Reader Enquiry No. 103

ES Industries has introduced its new line of Chromegabond Fluorosep-RP **HPLC columns**, designed for the separation of samples not resolvable with traditional reversed phase columns. The manufacturer states that high-resolution performance and durability over a wide range of pH and temperatures are consistently obtained with all these columns. The FSP (phenyl) columns are used for the separation of closely related geometric isomers, particularly those of aromatic compounds. FO (octyl) columns use a fluorocarbon phase for the separation of many pharmaceutical compounds and formulations. For the separation of proteins, peptides and

other related compounds of medical and biological interest, the company offers FP (propyl), a high-efficiency, short-chain phase. Reserved lots are also available to maintain maximum reproducibility for quality control applications.

Reader Enquiry No. 104

Bangs Laboratories has announced the availability of a new class of **microspheres** with 'easy binding' surface chemistries. The surface epoxy groups of these pre-activated microspheres are said to permit easy covalent coupling of antigens, antibodies, haptens or oligonucleotides. Any ligand with primary amino or other similarly reactive groups can be bound easily and quickly. Sizes range from about 30 nm to >6 µm, these microspheres are available as methylmethacrylate/epoxy or styrene/epoxy copolymers, with chloromethyl and aldehyde surface groups also offered. There are over 1,000 microspheres to choose from, including magnetic particles with other materials, sizes, colours, fluorescent derivatives and surface chemistries to choose from — protein-coated microspheres are also available.

Reader Enquiry No. 105



Bang's new epoxy microspheres.

LC and GC

Perkin-Elmer Sciex has introduced two new atmospheric-pressure ionization (API) benchtop **liquid chromatography/mass spectrometry (LC/MS)** systems — the API 100 (single quadrupole) LC/MS and the API 300 (triple quadrupole) LC/MS/MS. The API 300 allows all modes of MS/MS operation for full characterization of biopharmaceutical compounds and the specificity needed for new drug development. The API 100 enables users to obtain structural information using collisionally induced dissociation (CID/MS) together with a unique feature enabling MS and MS/MS to be carried out sequentially during the same analysis. For peptides and proteins, both systems allow

molecular weights to be determined with accuracies better than 0.01 per cent at 200K. BioToolBox software is available for Power Macintosh computers with both systems and enables protein and peptide characterization using CID/MS or MS/MS data.
Reader Enquiry No. 106

Varian states that its new **capillary-GC injector** can deliver up to a four-fold increase in analyte recoveries, allowing the analysis of even the most polar and thermally labile compounds. The 1078 universal capillary injector is said to decrease injector maintenance time by up to 50 per cent. It can extend the benefits of the company's septum-equipped programmable injector, to include split, splitless and temperature-ramped splitless injection for maximum injector flexibility. The injector also facilitates large-volume injections, allowing low parts-per-trillion detection levels by GC and GC/MS. Compatible with 50- to 530- μ m columns and capable of isothermal or temperature-programmed operation, the injector includes a new sealing design for easy removal and replacement of the insert.
Reader Enquiry No. 107

Gold Nouveau is a new HPLC product produced by Beckman, which includes a new **electronic autosampler and a new low-cost system control centre**. The system has optional bioinert (PEEK) μ -flow paths and improved flow electronics for superior performance in μ -bore chromatography. Users have a choice of front panel control or a Windows-based PC. A single-piston pump mechanism provides consistency and is said to be virtually unaffected by air bubbles, eliminating the need for solvent degassing. Detection is accomplished by a programmable, high-sensitivity, variable-wavelength ultraviolet detector or a high-resolution, diode-array detector, featuring real-time, peak-purity determination.
Reader Enquiry No. 108

The Micro-flow HPLC system from LabLogic offers lower flow rates and less mobile phase, yet faster chromatography. The company states that this results in improved resolution with its greater peak-to-valley ratios. For radioactive measurement, micro-HPLC may prove to be useful — narrow peaks mean smaller backgrounds and almost no waste disposal. Features include liquid-scintillation cells of 25 μ l or 50 μ l a scintillator pump that can deliver accurately 50 to 1,000 μ l min^{-1} , and solid scintillator cells with 0.7 μ l void volumes. The manufacturer states that a typical 30-min run, flowing at 1.0 ml min^{-1} can become a 10-min run with micro-HPLC at 50 μ l min^{-1} . With a 50:1 scintillator to eluate ratio, scintillator consumption can be reduced from 150 ml per run to 25 ml per run.
Reader Enquiry No. 109

Literature

The new 1996 **chromatography products catalogue**, from Supelco, features more information and products than ever before, making it an essential reference tool for the lab or office. Each product is backed by the company's expert technical service. There are over 1,000 new products offered, including more than 13,000 total products for GC, HPLC, LPLC, TLC, sample handling and chemical standards. There are 960 easy-to-follow pages, more than 500 application chromatograms and conditions, four indices to searched, and concise theory introductions and technical product descriptions.
Reader Enquiry No. 110

Sigma Chemical now offers over 500 **affinity resins** that can be applied to almost any purification need. This diverse selection of products is presented in a comprehensive brochure, organized by functional groups, such as hydrophobic, dye and nucleic acid resins.



Sigma's affinity chromatography catalogue.

Reader Enquiry No. 111

Hypersil has produced a **technical guide on the use of media for preparative- and process-scale chromatography**. This booklet includes an evaluation of slurry packing conditions, an investigation into the mechanical stability of a range of materials and the retention capacity and selectivity of a broad range of analytes. Several silica based C18 materials are considered and there are reprints of posters presented at meetings. The information supplied demonstrates the mechanical and chemical stability of HS BDS Hyperprep and its broad applicability to neutral, polar and basic compounds.
Reader Enquiry No. 112

Software selection

Russell Mainstream Supply has begun shipping Chrom Perfect/5890 Direct, the newest version of this **Windows-based chromatography data system**. The software package takes data directly from up to eight Hewlett-Packard 5890 gas chromatographs. It also controls set points on up to eight HP5890s, each of which can have two detectors plus an HP7673 autosampler. Both single- and dual-tower autosamplers are supported. All GC temperature and pressure set points can be changed in single and autosampler injection modes. GCs can be located up to 50 metres away from the computer and users can log data to a network file server for review by other users on the network.

It can acquire data from non-HP chromatograms through 24-bit A/D boards or PE-Nelson interfaces to provide a common environment for GC, HPLC, SEC and CZE applications.

Reader Enquiry No. 113

Aston Scientific, distributors of the popular EZChrom **chromatography data system**, now offer optional system suitability software. The software allows the user to select from 25 different fixed parameters plus five custom parameters, which can be used to test whether the chromatographic system is suitable for certain analyses. Parameters such as peak resolution, asymmetry and theoretical plates can be selected for report, with results compared to user-entered minimum and maximum values for each analyte. This is a Windows-based application that can be used with as many as four chromatographs with four detectors, each chromatograph operating completely independently of the others.

Reader Enquiry No. 114 □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Gramsch Laboratories

- New: Somatostat. Rec. 2A/B Antisera
 - G-Protein Antisera - α , β , γ -subunit
 - G-Protein Peptide Fragments
 - Opioid Receptor Antisera
 - Monoclonal Opioid-Antibody 3E7
 - YGGF... - specific; appl.: combinatorial peptide library tests
 - Cyclic Nucleotide Antisera
 - Antibody/Peptide Production Service
- Phone 08138 1769 or 08131 5679-0
Fax 08138 1260 or 08131 5679-24
WWW:ourworld.com/serve.com/
homepages/jgramsch/gli.htm
- Gramsch GmbH
D-85247 SCHWABHAUSEN, Germany

READER ENQUIRY NO 16

BIANORM[®] EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

READER ENQUIRY NO 15